Meik Dörpinghaus

On the Achievable Rate of Stationary Fading Channels

Springer

Foundations in Signal Processing, Communications and Networking

Series Editors: W. Utschick, H. Boche, R. Mathar

Meik Dörpinghaus

On the Achievable Rate
of Stationary Fading Channels

Springer

Series Editors:

Wolfgang Utschick
TU Munich
Associate Institute for Signal
Processing
Arcisstrasse 21
80290 Munich, Germany

Holger Boche
TU Berlin
Dept. of Telecommunication Systems
Heinrich-Hertz-Chair for Mobile
Communications
Einsteinufer 25
10587 Berlin, Germany

Rudolf Mathar
RWTH Aachen University
Institute of Theoretical
Information Technology
52056 Aachen, Germany

Author:
Meik Dörpinghaus
RWTH Aachen University
Institute for Theoretical
Information Technology
Templergraben 55
52056 Aachen
Germany
doerpinghaus@ti.rwth-aachen.de

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

To my parents and Anke

Preface

In typical mobile communication systems transmission takes place over a time-varying fading channel. The stochastic channel fading process can assumed to be bandlimited and its realization is usually unknown to the receiver. To allow for a coherent signal detection, the channel fading process is often estimated based on pilot symbols which are periodically inserted into the transmit symbols sequence. The achievable data rate with this approach depends on the dynamics of the channel fading process. For this conventional approach, i.e., performing channel estimation solely based on pilot symbols and using it for coherent detection (*synchronized detection*) in a second step, bounds on the achievable data rate are known. However, in recent years receiver structures got into the focus of research, where the channel estimation is iteratively enhanced based on the reliability information on data symbols (code-aided channel estimation). For this kind of systems, the bounds on the achievable data rate with synchronized detection based on a solely pilot based channel estimation are no longer valid.

The study of the possible performance gain when using such receivers with synchronized detection and a code-aided channel estimation in comparison to synchronized detection in combination with a solely pilot based channel estimation poses also the question on the capacity of stationary fading channels. Although such channels are typical for many practical mobile communication systems, already for the simple case of a Rayleigh flat-fading channel the capacity and the capacity-achieving input distribution are unknown. There exist bounds on the capacity, however, most of them are tight only in a limited SNR regime and rely on a peak power constraint.

Thinking of this, in the present thesis various aspects regarding the capacity/achievable data rate of stationary Rayleigh fading channels are treated. First, bounds on the achievable data rate with i.i.d. zero-mean proper Gaussian input symbols, which are capacity achieving in the coherent case, i.e., in case of perfect channel knowledge at the receiver, are derived. These bounds are tight in the sense that the difference between the upper and the lower bound is bounded for all SNRs. The lower bound converges to the coher-

ent capacity for asymptotically small channel dynamics. Furthermore, these bounds are extended to the case of multiple-input multiple-output (MIMO) channels and to the case of frequency selective channels.

The comparison of these bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to the achievable rate while using receivers with synchronized detection based on a solely pilot based channel estimation already gives an indication on the performance of such conventional receiver structures. However, for systems with receivers based on iterative code-aided channel estimation periodic pilot symbols are still used. Therefore, in a further part of the present work the achievable rate with receivers based on synchronized detection and a code-aided channel estimation is studied. For a specific type of such a receiver an approximate upper bound on the achievable rate is derived. The comparison of this approximate upper bound and the achievable data rate with receivers using synchronized detection based on a solely pilot based channel estimation gives an approximate upper bound on the possible gain by using this kind of code-aided channel estimation in comparison to the conventional receiver using a solely pilot based channel estimation. In addition, the achievable data rate with an optimal joint processing of pilot and data symbols is studied and a lower bound on the achievable rate for this case is derived. In this context, it is also shown which part of the mutual information of the transmitter and the receiver is discarded when using the conventional receiver with synchronized detection based on a solely pilot based channel estimation.

Concerning the typically applied periodic pilot symbols the question arises if these periodic pilot symbols are optimal from an information theoretic perspective. To address this question, the mutual information between transmitter and receiver is studied for a given discrete signaling set. The optimum input distribution, i.e., the one that maximizes the mutual information when restricting to the given signaling set, is given implicitly based on the Kullback-Leibler distance. Thereon it is shown that periodic pilot symbols are not capacity-achieving in general. However, for practical systems they allow for receivers with small computational complexity.

Acknowledgements

work was initiated by their intriguing question 'What can we gain by iteratively enhancing the channel estimation using reliability information on data symbols?'. Prof. Meyr and Prof. Ascheid gave me the freedom and time to follow my ideas to use information theory to find answers to the given question. I want to thank them for their support over the years. Their continuous encouragement and faith in my abilities have been very motivating. Especially, I would like to thank Prof. Meyr for very valuable discussions which have been particularly fruitful by bringing together his deep knowledge on estimation and detection theory and my thoughts and ideas in information theory.

Furthermore, I thank Prof. Helmut Bölcskei for reading an earlier version of the present manuscript. In addition, I appreciate that he invited me to the Communication Theory Group at ETH Zurich in June 2007. This exciting stay in Zurich gave me invaluable insights and the opportunity to discuss with Giuseppe Durisi, Ulrich Schuster, and Veniamin Morgenshtern. Besides the fact that all of them are outstanding researchers, I would like to thank them for their friendly welcome and their openness in sharing their knowledge. With Veniamin I have discussed my approach of bounding the achievable rate of a stationary Rayleigh flat-fading channel using means of random matrix theory. Although we did not came up with a solution, this collaboration has been rewarding as we understand the underlying problems. During these discussions I learned a lot from Veniamin who has a deep knowledge in random matrix theory. Furthermore, I have to thank Giuseppe and Uli for discussing several aspects of the capacity of noncoherent fading channels during my stay at ETH. In addition, I gracefully thank Giuseppe Durisi, with whom I discussed large parts of the present work when he came to Aachen in June 2009. By pinpointing weaknesses in some of my proofs, his comments and critics have enhanced the present work a lot.

I also thank Prof. Rudolf Mathar for interesting discussions on the parts regarding discrete input distributions covered in Chapter 10. This part of the work emerged after a talk on *Capacity-Achieving Discrete Signaling over Additive Noise Channels* given by him on the UMIC day in 2007. A few weeks later I discussed the application of this approach to the scenario of a stationary Rayleigh flat-fading channel with him, resulting in an ISIT publication.

In addition, I thank my colleagues at the Institute for Integrated Signal Processing Systems for a pleasant working environment. Especially I would like to thank Adrian Ispas, Lars Schmitt, Susanne Godtmann, Martin Senst, and Dan Zhang for many helpful and inspiring discussions. Special thanks are due to Adrian Ispas with whom I had endless discussions about the material in the present manuscript especially during the final phase of the work.

Last but not least, I thank my parents for their continuous support during my studies enabling all this. Lastly, I am particularly indebted to my girlfriend Anke, for her patience and her encouragement which essentially contributes to the success of this work.

Aachen, August 2010   *Meik Dörpinghaus*

Contents

Chapter 1
Introduction

Over the last decade the wireless communication market evolved dynamically, enabling a variety of new services and applications beside the classical voice telephony. One example is the mobile internet. This development was enabled by enhanced technology and rapidly increasing available data rates. While the demand for higher data rates stimulates the development of systems providing increased bandwidth, the available frequency spectrum is limited as a natural resource, requiring that deployed technology uses this resource with increasing efficiency. Therefore, one of the major goals of telecommunication research is the increase of *spectral efficiency* leading to strong efforts to enhance physical layer design. However, physics of the wireless channel pose strict limits on the spectral efficiency. To evaluate the performance of systems in operation it is highly beneficial to know about the ultimate physical limits. In addition, knowledge on the limits in capacity of the physical communication channel is a valuable basis on decisions and guidelines for research and development of physical layer designs. If existing systems show to operate already close to these ultimate limits, the revenue of putting additional effort in enhancing existing technology will be very small. Therefore, a reliable judgment on research perspectives premises a detailed comprehension on the theoretical limits.

In the context of communication technology, the discipline of *information theory*, aims to give answers regarding the limits imposed by the physical channel. Historically, information theory originated from work on the fundamental limits on reliable communication and compression of data by Claude E. Shannon. In his landmark paper [108], Shannon identified the channel *capacity* as the maximum mutual information between the input and the output of a channel. He stated the channel coding theorem, in which for *discrete memoryless channels* (DMC) Shannon proved that all rates below channel capacity are achievable. In addition, for the additive white Gaussian noise (AWGN) channel with an average input power constraint Shannon was able to quantify the channel capacity exactly, as well as to determine the capacity-achieving input distribution to be Gaussian.

Since Shannon's days, information theory delivered a variety of valuable results helping to comprehend the theoretical background of modern communication systems. However, many problems are still open. In addition to the previously mentioned AWGN channel, in mobile wireless communication environments the signal is additionally disturbed by a multiplicative random process which is characterized by a temporal correlation. Therefore, this channel is not memoryless. The realization of the channel fading process is typically unknown to both, the transmitter and the receiver. Although this class of channels applies to many realistic mobile communication systems, the question on its capacity and on the capacity-achieving input distribution is still open.

In many typical receivers, estimates of the channel fading process are calculated to enable coherent detection[1]. For this purpose, typically pilot symbols, i.e., symbols which are known to the receiver, are inserted into the transmit data stream. Due to the temporal correlation of the channel fading process, they allow to estimate the fading process by interpolation. The principle of using the estimate of the channel fading process as the actual fading process for coherent detection is well known and is often referred to as *synchronized detection* [79]. The use of such training sequences can be understood as a specific type of coding [67]. However, their optimality with respect to the channel capacity has never been shown. Concerning the system design, they have the advantage to allow for a separation of synchronization and decoding within the receiver. The aim of the so called *inner receiver* is to estimate the unknown channel parameters, while the so called *outer receiver* performs decoding of the transmitted information sequence, see Fig. 1.1. This separation enables receiver implementations with suitable complexity. While in conventional designs of a synchronized detection based receiver the channel estimation is solely based on pilot symbols, recently iterative receiver concepts have been proposed, see, e.g., the special issue [58], the editorials [110], [111], and the publications [83] and [98]. Corresponding to these iterative receiver concepts, synchronization, respectively channel estimation, is enhanced using reliability information on the transmitted data symbols delivered by the channel decoder, as shown in Fig. 1.1 in light gray. Regarding the channel estimation often the term *code-aided* channel estimation is used in this context. We refer to this receiver structure as *iterative code-aided synchronized detection* or *iterative synchronization and decoding*. The iterative concatenation of synchronization and decoding breaks up the strict separation between inner and outer receiver.

While for the conventional approach, where synchronized detection is used in combination with a solely pilot based channel estimation, tight[2] upper and lower bounds on the achievable data rate are known, see [6], [5], and

---

[1] We do not consider differential modulation as their performance is known to be worse with respect to the achievable rate.

[2] The term *tight* means that the bounds are sufficiently close from an engineering point of view and is not meant in the sense of mathematical tightness.
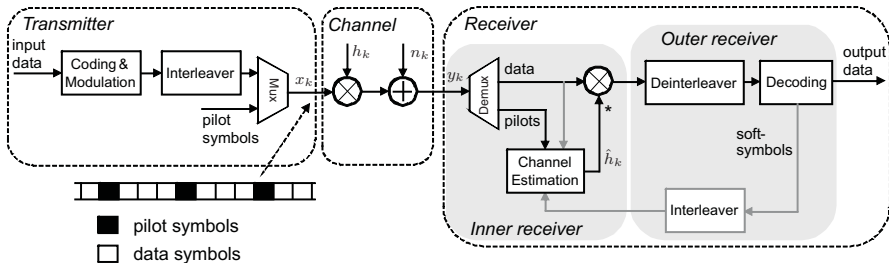
Fig. 1.1: **Block diagram of a transmission system with a receiver based on synchronized detection**, exemplary for BPSK and a single-input single-output flat-fading channel in a simplified discrete-time baseband representation; iterative code-aided channel estimation in light gray; $\{x_k\}$ transmit sequence, $\{h_k\}$ channel fading process, $\{\hat{h}_k\}$ estimate of channel fading process, $\{n_k\}$ additive noise process, $\{y_k\}$ received sequence

[77], for receivers based on iterative code-aided synchronized detection, these bounds are not valid. Besides the general interest in an answer to the capacity of fading channels this motivates to study the achievable rate without any assumption on the use of pilot symbols, i.e., a specific type of coding.

The aim of this thesis is the study of bounds on the achievable rate/ capacity[3] of stationary fading channels where the channel state information (CSI) is unknown to the transmitter and the receiver, while at least the receiver is aware of the channel law. This capacity is sometimes referred to as *noncoherent capacity*.

Beside the fact that a characterization of the channel capacity would enable the evaluation of the possible performance gain when using code-aided instead of solely pilot based synchronization, it would also help to answer several other questions in communication system design. Concerning multiantenna (MIMO) systems and wideband channels, knowledge of the capacity and its dependency on the number of transmit antennas, the spatial antenna correlation and the frequency selectivity of the channel is very important, as the number of degrees of freedom of the channel increases with the number of antennas and the bandwidth of the transmitted signal. Thus, spreading the transmit power over too many transmit antennas or a too large bandwidth, leads to a degradation of the achievable rate.

---

[3] Depending on context we use the term *achievable rate* instead of the term *capacity* when we make presumptions on the input distribution, e.g., assuming identically distributed (i.d.) or independent identically distributed (i.i.d.) input symbols, which are not capacity-achieving in general.

1.1 Prior Work

In the following, we give a brief overview on the existing literature regarding the capacity of fading channels. This presentation especially focuses on the specific assumptions that have been made to simplify the problem. Therefore, it has not the aim to be an exhaustive list of existing work.

The capacity of fading channels where the channel state information is unknown has received a lot of attention in the literature. One line of work considers the simple block fading channel, where the channel is assumed to be constant over a block of N symbols and changes independently from block to block. In [75] the capacity of a MIMO block fading channel is studied. Furthermore, results concerning the optimal input distributions in case of a low SNR noncoherent MIMO block fading channel have been recently presented in [115]. The achievable rate for a MIMO block fading channel when using training sequences for channel estimation has been studied in [48], [94] and [37]. However, the block fading model is nonstationary and therefore different from the scenario we consider in the present work.

In contrast, in [6], [5], [77], and [55], the achievable rate of time-continuous fading channels has been examined under the assumption that estimates of the channel are available. In [77] a general setting has been studied, whereas in [6], [5], and [55] the channel estimates are acquired by a solely pilot based channel estimation. Furthermore, in [22] the achievable rate with solely pilot based synchronized detection has been compared to a lower bound on the achievable rate with independent identically distributed (i.i.d.) Gaussian signaling. Opposed to these works, where the channel estimation is solely based on pilot symbols, recently, also a lower bound on the achievable rate with joint processing of pilot and data symbols has been given in [56] in the context of a block fading channel.

Over the last years, the capacity of time-continuous fading channels without any assumption on pilot symbols has been the subject of various contributions. However, this problem turns out to be notoriously difficult. Even for the memoryless case there is no closed form solution for the capacity [119], [1]. In [68] Lapidoth and Moser introduced a quantity called *fading number* as the second-order term in the high signal-to-noise ratio (SNR) asymptotic expansion of capacity, to characterize the effect of the unknown channel depending on its statistics. Based on the fading number, in [68], [66], and [67] the asymptotic high SNR capacity of a stationary Gaussian flat-fading channel has been investigated. This examination shows that the high SNR behavior significantly depends on the channel model. If the channel fading process is *non-regular*, its high SNR slope shows a pre-log behavior, whereas *regular* fading yields a log log behavior for high SNR.[4] The approach based on the

---

[4] A *non-regular* fading process is characterized by a channel prediction error variance converging to zero for infinite SNR. Otherwise the fading process is *regular*, see also [24] and Chapter 2.

fading number has also been used in the further work in [69], [59], and [81] for a study of the fading number in the MISO case, the derivation of non-asymptotic upper bounds on capacity, extensions to the MIMO case, and the characterization of input distributions that achieve the fading number. All of the works based on the fading number mainly focus on the high SNR regime.

In contrast to this, in [34] an approximate behavior of the capacity for different SNR regimes depending on the channel prediction error, which is directly related to the channel dynamics, has been considered. This work mostly considers fading channels, which are characterized by a first order Gauss-Markov model, which shows, due to the fact that it is *regular*, a log log high SNR behavior of capacity.

The following line of work mainly focuses on the low SNR case. In [105] lower bounds on capacity for peak power constrained input signals are given. Furthermore, in [104], [106], and [107] also upper bounds on the capacity have been derived and analyzed with special emphasizes on the low SNR regime, including extensions to the MIMO and the frequency selective case. Furthermore, also [133] focuses on the low SNR case when discussing the achievable rate for the special case of PSK signaling.

Such a restriction to PSK input symbols has also been made in [97] for the evaluation of the system capacity of wideband (OFDM) communication. Furthermore, the peak power constrained capacity for the more general case of underspread frequency-selective stationary fading channels has been discussed in [30], [31], [33], and [99]. Therefor, in [33] a detailed derivation of a discrete-time discrete-frequency system model outgoing from the genuine continuous-time model is shown. In [101] and [102] this work has been extended to the MIMO scenario. While the results in these contributions rely on a strictly limited support of the channel's scattering function, in [32] the sensitivity of the capacity on this channel model has been evaluated. Such an evaluation is important as it has been shown in [60], [62], [61], and [63] that the high SNR capacity of a frequency-selective fading channel strongly depends on the details of the considered channel model. I.e., the capacity grows for asymptotically high SNR bounded or unbounded, depending on the number of delay paths an their gains.

So far, we discussed prior work on the one hand for the block fading channel, and on the other hand for the stationary fading channel. A block-stationary channel is a combination of a stationary and a block fading channel. Within a block it behaves like the stationary channel with temporal correlation. The blocks themselves might be correlated between each other or might be independent. The special case where the fading between the blocks is independent is typically referred to as time-selective block fading channel. The capacity of such a block-stationary channel has been studied in [71] and in [11]. These publications mainly focus on the behavior of capacity with respect to the SNR, the block length, and in [71] the number of antennas.

To gain some more fundamental insight into the behavior of communication over noncoherent fading channels, for the specific case of a MIMO

block fading channel, in [134] a capacity expression has been given, which has the geometric interpretation as sphere packing on the Grassmann manifold. Furthermore, [70] poses the question 'how perfect need "perfect side information" be?' and discusses the robustness of nearest neighbor decoding, which results from the assumption of perfect channel knowledge, with respect to estimation errors of the channel. An extension of this work to the MIMO case has been presented in [131].

A further approach to simplify the study on the capacity is to fix the input distribution. In this context, in [12] the achievable rate for the special case of Gaussian inputs has been discussed for a Gauss-Markov fading channel. Such a restriction to Gaussian input symbols will also be used at various places within the present work to calculate the achievable rate with this kind of input distribution for stationary fading with a compactly supported power spectral density.

Remark: Notice, for clarity of presentation we have not always explicitly stated, if a publication only considers the SISO case or also the MIMO scenario.

1.2 Objectives and Contributions

The general objective of the present work is to get an enhanced comprehension of the achievable rate of stationary Rayleigh fading channels. So on the one hand, this study should deliver enhanced bounds on the achievable rate over this class of channels. On the other hand, it should establish the link to the bounds on the achievable rate with synchronized detection and, thus, enable an understanding of the achievable rate with synchronized detection based receivers including iterative synchronization.

In contrast to many existing bounds on capacity, which focus on the asymptotic SNR behavior, e.g., [67] for the high SNR regime and, e.g., [107] for the low SNR regime, our aim is to get bounds on the achievable rate that are useful over a wide range of the SNR, i.e., the bounds should not be such loose that they give no insight. Most of the time, we will use the term *achievable rate* instead of *capacity*, as in the majority of the following work, we make restrictions on the input distribution, e.g., assuming identically distributed (i.d.) or i.i.d. inputs. First, these assumptions on the input distributions simplify the mathematical treatment of the problem and will turn out to perform well for a wide range of channel parameters. Secondly, in case the channel realization is known to the receiver, i.e., the coherent scenario, the capacity-achieving input distribution is i.i.d. zero-mean proper Gaussian. This already leads to the next demand we have on bounds on the achievable rate. In addition, different to a lot of the existing work, we are interested in bounds on the achievable rate which can also be used in combination with proper Gaussian input distributions, i.e., without a peak power constraint. We have two

motivations to study the achievable rate with i.i.d. zero-mean proper Gaussian inputs: On the one hand, they are capacity-achieving in case of perfect channel knowledge. Although it is well known that they are not optimal for unknown channel state information at the receiver [70], in [12] bounds on the mutual information with Gaussian input distributions have been derived for a Gauss-Markov fading channel[5], showing that at moderate SNR and/or slow fading, Gaussian inputs still work well. On the other hand, many known upper bounds on the capacity based on a peak power constraint get loose for high peak-to-average signal powers. Furthermore, when restricting to peak power constrained input symbols, the achievable rate will not converge to the coherent capacity for asymptotically small channel dynamics. In addition, in many cases the capacity-achieving input distribution becomes peaky and, thus, impractical for real system design. In contrast, proper Gaussian input distributions serve well to upper-bound the achievable rate with practical modulation and coding schemes, see also [93]. Therefore, it is interesting to see, how the achievable rate behaves for i.i.d. zero-mean proper Gaussian inputs, which are not peak power constrained. In conclusion, our motivation is to get bounds which give much insight into the behavior of the achievable rate over a wide range of the SNR, while we accept for the fact that these bounds are not longer bounds on capacity when we fix the input distribution.

In the following, we give a detailed overview regarding the subjects studied in this thesis. In the present work, first, we investigate the achievable rate of a stationary discrete-time Rayleigh flat-fading channel. Its realization is unknown to both the transmitter and the receiver, while the receiver is aware of the channel law. In addition, we assume that the power spectral density (PSD) of the fading process has compact support. Furthermore, we assume the support region of the PSD of the channel fading process to be smaller than the inverse of the symbol duration, i.e, we consider *nonregular* fading [24].

Under these assumptions, we derive different bounds on the achievable rate. First, we derive a set of bounds which is based on a pure mathematical evaluation of the achievable rate. For the derivation of the upper bound, we need the restriction to i.d. input distributions. Secondly, we derive an alternative upper bound on the achievable rate, which is based on the channel prediction error variance. For the derivation of this bound, we assume i.i.d. input symbols. On the one hand, we evaluate the given bounds for general i.d. respectively i.i.d. input distributions with an average and a peak power constraint. On the other hand, we also give the bounds for the special case of i.i.d. zero-mean proper Gaussian input symbols. Therefore, it is important to note that the derivation of our upper bounds does not rely on a peak power constraint. This is a major difference to many other known bounds, e.g., in [67] and [107]. The derived bounds depend on the channel dynamics via

---

[5] Notice that in case of Gauss-Markov fading, the PSD of the channel fading process is characterized by an unbounded support. In contrast to this, within the present work, we assume a PSD of the channel fading process with bounded support, see Chapter 2.

the PSD of the channel fading process. Some of them hold for general power spectral densities while others only hold for the special case of a rectangular PSD. We compare the bounds on the achievable rate also to known bounds on the capacity given in [67] and [107], and to bounds on the achievable rate when using synchronized detection in combination with a solely pilot based channel estimation as studied in [6].

Besides the general bounds on the achievable rate with i.d./i.i.d. input distributions, we also study the achievable rate with a receiver using synchronized detection in combination with iterative code-aided channel estimation, as it has been previously described, see Fig. 1.1. On the one hand, we clearly identify the nature of the possible gain by iteratively enhancing the channel estimation based on reliability information on the data symbols in comparison to a solely pilot based channel estimation. The possible gain arises due to the information contained in the temporal correlation of the channel estimation error of a solely pilot based channel estimation. This information is discarded when using synchronized detection as the detector, which uses a symbol-wise detection metric, cannot exploit the information contained in the temporal correlation of the channel estimation error. Although a receiver using synchronized detection in combination with an iterative code-aided channel estimation still uses a symbol-wise detection metric, parts of the information contained in the temporal correlation of the channel estimation error when using a solely pilot based channel estimation can be exploited by such an enhanced receiver structure.

In addition to this fundamental understanding of the possible gains, we give an upper bound on the achievable rate for a specific receiver structure using synchronized detection with an iterative code-aided channel estimation. The widely known and studied receiver using synchronized detection based on an iterative code-aided channel estimation can be deduced by expressing the common joint ML detection and MAP parameter estimation problem [79] based on a set of fixed point equations [98], or alternatively based on the expectation maximization (EM) framework [40]. In contrast, we calculate an upper bound on the achievable rate for a slightly modified receiver. The modification lies in the channel estimator which in contrast to the channel estimator given in [98] does not use the observation y_k for the estimation of the fading weight h_k. This modification is required for mathematical reasons. However, we guess that for small channel dynamics the amount of discarded information is very small and therefore, this modification is not a severe restriction. Furthermore, we assume i.i.d. zero-mean proper Gaussian data symbols. We show which part of the mutual information between the transmitter and the receiver can be exploited by the given receiver structure. I.e., we show that there exists a component of the mutual information, which by the combination of symbol-wise detection and the channel estimator which does not use the observation y_k for the estimation of h_k cannot be retrieved. However, for small channel dynamics, this amount of information seems to be small. The question if and which part of this information can be retrieved

when using a channel estimator without the described modification remains open. The given upper bound on the achievable rate for the studied receiver still explicitly depends on the channel interpolation error variance at an arbitrary time, which itself is a function of all past and all future data symbols. Unfortunately, we have not found a closed form solution for an upper bound. But, at least for small channel dynamics it is reasonable to approximate the channel interpolation error variance by the interpolation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols. This leads to an approximative upper bound which is valuable, as it gives a tighter approximative upper bound on the achievable rate with a practical receiver structure than the upper bound on the achievable rate with i.i.d. input symbols.

As already mentioned before, also receivers using synchronized detection and an iterative code-aided channel estimation, still use a symbol-wise detection metric. However, we show that such a symbol-wise detection metric is in general not optimal. Therefore, we also study the achievable rate when performing an optimal joint processing of pilot and data symbols. For this case we give a lower bound on the achievable rate for the case of a stationary Rayleigh flat-fading channel. This lower bound, can be seen as an extension of a lower bound on the achievable rate with joint processing of pilot and data symbols for the case of a block fading channel recently given in [56].

Afterwards, we extend the bounds on the achievable rate with i.d. input symbols, i.e., again considering no pilot symbols, given for the case of a single-input single-output (SISO) Rayleigh flat-fading channel, based on a pure mathematical derivation, to the case of a multiple-input multiple-output (MIMO) system considering also spatial antenna correlation. As for the SISO case, we compare the bounds on the achievable rate for i.i.d. zero-mean proper Gaussian input symbols with bounds on the achievable rate when using synchronized detection in combination with a solely pilot based channel estimation as studied in [5].

In addition, we extend the bounds on the achievable rate with i.d. input symbols, which we derive for the SISO flat-fading channel, to the case of an underspread frequency-selective (wideband) channel with a compactly supported scattering function. The derivation is based on a discrete-time discrete-frequency system model derived in [33] for underspread wide-sense stationary uncorrelated scattering. The main novelty in this context is the derivation of bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. Especially the derivation of the upper bound is new as it is not based on a peak power constraint. In contrast, the bounds on the capacity given in [33] explicitly require a peak power constraint. However, the given upper bound for proper Gaussian input symbols holds only for the special case of a brick-shaped scattering function and holds only approximately for finite bandwidth. Also for the frequency-selective case, we compare the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input

symbols to the achievable rate using synchronized detection in combination with a solely pilot based channel estimation discussed in [4] and [47].

In the further discussion, we again deviate from the assumption on i.d. or i.i.d. input symbols, which are not optimal in general [107]. For the case of a block-stationary SISO Rayleigh flat-fading channel with independent faded blocks, i.e., a time-selective block fading channel, we study the optimum distribution over a given set of input sequences. Thus, we restrict to discrete input distributions and drop the assumption on i.i.d. input symbols. There are two main objectives to study discrete input distributions. First, in [10] it has been shown that discrete input distributions are capacity-achieving for a variety of conditionally Gaussian channels with bounded input constraints. E.g., for the case of a Rayleigh flat-fading channel without temporal correlation, it has been shown that the capacity-achieving input distribution is discrete with a finite number of mass points [1]. The scenario in the present work falls into the class of conditionally Gaussian channels. These observations and secondly the fact that practical realizable systems use discrete input distributions are the motivation to study the *constrained capacity*[6] for the given scenario with the restriction to discrete input distributions. We show that the optimum input distribution over a fixed set of given signaling sequences is characterized by a constant Kullback-Leibler distance between the output probability density functions (PDF) conditioned on the individual input sequences and the unconditional output PDF. For the special case of PSK signaling, we explicitly characterize the set of optimum input distributions, which corresponds to a uniform distribution over transmit sequences that can be distinguished by the receiver. It is shown that the special case of transmitting one pilot symbol per fading block, i.e., a symbol that is known to the receiver, and independent input symbols at all other time instances lies within this set and, thus, is optimum. In addition, the asymptotic high SNR constrained capacity is degraded by at least a factor of $\frac{N-1}{N}$ with respect to the case of perfect channel state information at the receiver when considering time-selective block fading with a block length of N. The result that the use of one pilot symbol per block and i.i.d. data symbols at all other time instances for PSK sequences is optimum allows us to conclude that in the context of stationary fading the use of pilot symbols which are periodically inserted into the transmit sequence is not optimal from a capacity point of view, at least when using PSK modulation. However, it should be mentioned that the use of periodic pilot symbols is a smart approach, as they allow for receivers of low computational complexity.

---

[6] Here we use the term *constrained capacity* instead of *achievable rate* as we optimize over the input distribution, which is only constrained concerning the set of signaling points.

Contributions

The specific contributions of the present work are summarized in the following:

- We calculate a new upper bound on the achievable rate with i.d. input symbols for the specific case of a rectangular PSD. The novelty of this bound lies in the derivation which is not based on any assumption of a peak power constraint. Therefore, we can evaluate this bound also for zero-mean proper Gaussian input symbols. For the case of a peak power constraint this upper bound is equivalent to an upper bound on capacity given in [107].
- With the upper bound on the achievable rate and a known lower bound on the capacity, we have found a set of bounds, which for zero-mean proper Gaussian inputs is tight in the sense that their difference is bounded. We are able to bound the gap analytically by $\gamma(1 + 2f_d)$ [nat/channel use], with the Euler constant $\gamma \approx 0.577$ and the maximum normalized Doppler frequency f_d. Furthermore, the upper bound is equal and the lower bound converges to the coherent capacity for asymptotically small channel dynamics.
- We show that the achievable rate with i.i.d. zero-mean proper Gaussian input symbols has the same high SNR slope, i.e., *pre-log*, as the channel capacity given in [67].
- We derive a new upper bound on the achievable rate with i.i.d. input symbols based on the channel prediction error variance. In contrast to existing bounds based on the channel prediction error variance, this bound is not restricted to peak power constrained input symbols. For the derivation of this upper bound we show that the calculation of the prediction error variance under the assumption that all past symbols are constant modulus symbols leads to a lower bound on the output entropy rate conditioned on the channel input, i.e., $h'(\mathbf{y}|\mathbf{x})$ and, thus, to an upper bound on the achievable rate. This upper bound holds for channel fading processes with an arbitrary PSD of compact support.
- We compare the different bounding approaches among each other and with the achievable rate based on a solely pilot based synchronized detection.
- We identify the nature of the possible gain that is achievable when using iterative synchronization and decoding in comparison to solely pilot based synchronized detection. This possible gain arises due to the information contained in the temporal correlation of the channel estimation error process. This information is discarded by performing synchronized detection in combination with a solely pilot based channel estimation as in this case detection is performed symbol-wise. Although receivers based on synchronized detection and an iterative code-aided channel estimation also use a symbol-wise detection metric, parts of the information contained in the

temporal correlation of the channel estimation error when using a solely
pilot based channel estimation can be exploited.

- We give an upper bound on the achievable data rate while using a specific
receiver based on synchronized detection and iterative code-aided channel
estimation, i.e., a receiver consisting of two dedicated units for channel
estimation and decoding. This bound is based on the assumption that
the data symbols are i.i.d. zero-mean proper Gaussian. Differently to the
typically studied receiver based on synchronized detection and iterative
code-aided channel estimation, which can be deduced from the joint ML
detection and MAP parameter estimation problem based on a set of fixed
point equations or by the EM framework, we study a receiver whose chan-
nel estimator does not use the observation y_k for the estimation of h_k. We
show that there is a part of the mutual information between the transmit-
ter and the receiver which cannot be exploited by this receiver structure.
However, for practical channel dynamics this part seems to be small. The
given upper bound on the achievable rate is a non-closed form expression.
It explicitly depends on the channel interpolation error variance for a spe-
cific but arbitrarily chosen time instant, which itself is a function on all
past and future transmit symbols. For small channel dynamics it is reason-
able to approximate this interpolation error variance by the interpolation
error variance calculated under the assumption that all past and future
transmit symbols are constant modulus symbols. However, we do not have
a proof that this yields a true upper bound. In contrast, for the special
case of constant modulus signaling we are able to give the upper bound
on the achievable rate with this iterative receiver structure in closed form.

- We give a lower bound on the achievable rate in case of an optimal joint
processing of pilot and data symbols for the case of a stationary Rayleigh
flat-fading channel. Here, we assume periodic pilot symbols, where the pilot
spacing is chosen such that the channel is sampled at least with Nyquist
rate. For the data symbols we only apply an average power constraint.
The given lower bound is achievable with i.i.d. zero-mean proper Gaussian
data symbols.

- We extend the bounds on the achievable rate with i.d. input symbols from
the SISO case to the MIMO scenario including spatial antenna correlation.
The main novelty of these bounds lies in the fact that the derivation of the
upper bound does not rely on a peak power constraint and thus enables
bounding of the achievable rate with i.i.d. zero-mean proper Gaussian in-
puts. Like in the SISO case, we compare the bounds on the achievable rate
with i.i.d. zero-mean proper Gaussian inputs to the achievable rate using
synchronized detection in combination with a solely pilot based channel
estimation for i.i.d. zero-mean proper Gaussian data symbols.

- Furthermore, we extend the bounds on the achievable rate with i.d. in-
puts to the case of underspread frequency-selective channels. Therefore,
we assume a brick-shaped scattering function. The main novelty lies in the
derivation of an approximate upper bound on the achievable rate which

holds also for non-peak power constrained inputs and, thus, enables the characterization of the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for frequency-selective fading channels depending on the delay and Doppler spread. Like in the flat-fading case, we compare the bounds on the achievable rate to the achievable rate with synchronized detection based on a solely pilot based channel estimation.

- For a time-selective block Rayleigh flat-fading channel and a fixed discrete set of input sequences, we show that the input distribution that maximizes the constrained capacity is characterized by a constant Kullback-Leibler distance between the output PDFs conditioned on the individual input sequences and the unconditional output PDF. For the special case of PSK signaling the optimum input distribution corresponds to a uniform input distribution over all distinguishable input sequences. Furthermore, the special case of using one specific time instant per fading block for a pilot symbol is included in the set of constrained capacity-achieving input distributions. The asymptotic high SNR constrained capacity is degraded at least by a factor of $\frac{N-1}{N}$ with respect to the case of perfect channel state information at the receiver, where N is the fading block length. These results also show that for stationary fading the use of periodically inserted pilot symbols is not optimal from a capacity point of view while using PSK modulation.

1.3 Outline

This thesis is structured as follows. In Chapter 2, we introduce the mathematical system model of the discrete-time SISO flat-fading channel, discuss the limitation of this system model and give a short recall of the meaning of operational and information theoretic capacity. Subsequently, we derive upper and lower bounds on the achievable rate with i.d. input symbols for a SISO system in Chapter 3, which are on the one hand evaluated for peak power constrained input symbols and on the other hand for i.i.d. zero-mean proper Gaussian symbols. In addition, we study the tightness of the upper and lower bound on the achievable rate and we examine the asymptotic high SNR behavior of the achievable rate with i.d. inputs and show that the achievable rate with i.i.d. zero-mean proper Gaussian inputs is characterized by the same high SNR slope as the channel capacity. Furthermore, we compare the given bounds on the achievable rate with i.d. input symbols with capacity bounds given in [107], [105], and [67]. Afterwards, in Chapter 4, an alternative upper bound on the achievable rate with i.i.d. input symbols is derived based on the channel prediction error variance. In Chapter 5, bounds on the achievable rate with solely pilot based synchronized detection are recalled and compared to the bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. Ensuing, in Chapter 6, we study the achievable rate with iterative

synchronization and decoding based receivers. Subsequently, in Chapter 7 the achievable rate with an optimal joint processing of pilot and data symbols is examined. In Chapter 8, the discrete-time flat-fading system model is extended to the case of multiple transmit and receive antennas and the bounds on the achievable rate given in Chapter 3 are extended to the MIMO case. Furthermore, we compare the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to the achievable rate while using synchronized detection in combination with a channel estimation solely based on pilot symbols. In Chapter 9, we extend the bounds on the achievable rate with i.d. input symbols to the case of a frequency-selective (wideband) channel. Furthermore, for comparison we recall bounds on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation. In Chapter 10, we study the constrained capacity for a time-selective block Rayleigh flat-fading channel for a given discrete signaling set. Finally, in Chapter 11, we summarize the work and give conclusions. The appendices contain several mathematical derivations and proofs required in the preceding chapters.

Chapter 2
Discrete-Time Flat-Fading System Model

In the present chapter, we introduce the discrete-time stationary Rayleigh flat-fading system model as it will be used throughout the following derivations. Hereby, we also discuss the limitations of the model and recall the definitions of operational and information theoretic capacity as a basis for our further derivations. In this chapter, we restrict to the single-input single-output (SISO) channel. In Chapter 8 the model is extended to the multiple-input multiple-output (MIMO) scenario, and in Chapter 9 the frequency-selective case is discussed.

Fig. 2.1 shows a basic block-diagram of a transmission system based on a flat-fading channel. This model is based on several simplifications as it only shows a discrete-time baseband representation as we will use in the following. This discrete-time representation corresponds to the symbol rate. Clearly, a symbol rate representation cannot be used to study timing- and frequency synchronization which would require a modeling with a sampling rate higher than the symbol rate to get a sufficient statistic. As timing- and frequency synchronization are out of scope of the present work, for the moment this model is of sufficient detail to show the main effects concerning capacity evaluations. In Section 2.3 we discuss the limitations of this model and in Section 9.1, we extend the model to the case of wide-sense stationary uncorrelated scattering (WSSUS) frequency-selective fading channels, starting from a continuous-time representation.
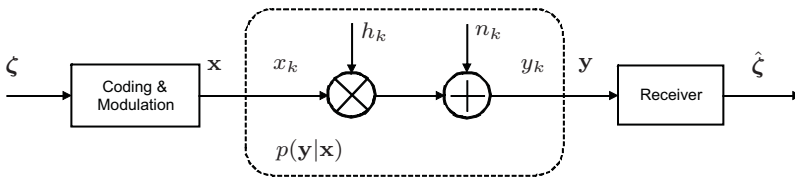
Fig. 2.1: Simplified block diagram of transmission system based on a discrete-time flat-fading channel

A binary data sequence $\boldsymbol{\zeta}$ of length N_{info} with the elements $\zeta_n \in \{0,1\}$ is mapped by encoding and modulation to a transmit symbol sequence of length N represented by the vector $\mathbf{x}$ with the elements x_k. The transmit symbols x_k are corrupted by multiplication with the channel fading weight h_k and by additive noise n_k. Here, it is important to note that the channel fading weights h_k are assumed to be temporally correlated and the process $\{h_k\}$ is stationary. With respect to the discrete-time system model shown in Fig. 2.1, the channel output observations y_k are contained in the vector $\mathbf{y}$. From a mathematical point of view, the effect of the channel is completely described by the conditional probability density function (PDF) $p(\mathbf{y}|\mathbf{x})$. The aim of the receiver is to find an estimate $\hat{\boldsymbol{\zeta}}$ of the binary data sequence $\boldsymbol{\zeta}$ based on the observation vector $\mathbf{y}$. As we are interested in the evaluation of the channel capacity/achievable rate, we do not further discuss how the receiver acquires $\hat{\boldsymbol{\zeta}}$ based on the observation $\mathbf{y}$.

After this brief overview of the system model, we now focus on an exact mathematical representation of the system model required for further study. We consider an ergodic discrete-time jointly proper Gaussian [82] flat-fading channel, whose output at time k is given by

$$y_k = h_k x_k + n_k \tag{2.1}$$

where $x_k \in \mathbb{C}$ is the complex-valued channel input, $h_k \in \mathbb{C}$ represents the channel fading coefficient, and $n_k \in \mathbb{C}$ is additive white Gaussian noise. The processes $\{h_k\}$, $\{x_k\}$, and $\{n_k\}$ are assumed to be mutually independent.

We assume that the noise $\{n_k\}$ is a sequence of i.i.d. proper Gaussian random variables of zero-mean and variance σ_n^2. The stationary channel fading process $\{h_k\}$ is zero-mean jointly proper Gaussian. In addition, the fading process is time-selective and characterized by its autocorrelation function

$$r_h(l) = \mathrm{E}[h_{k+l} h_k^*]. \tag{2.2}$$

Its variance is given by $r_h(0) = \sigma_h^2$.

The normalized PSD of the channel fading process is defined by

$$S_h(f) = \sum_{l=-\infty}^{\infty} r_h(l) e^{-j2\pi l f}, \quad |f| < 0.5 \tag{2.3}$$

where we assume that the PSD exists and $j = \sqrt{-1}$. Here the frequency f is normalized with respect to the symbol duration T_{Sym}. In the following, we will in general use the normalized PSD and, thus, refer to it as PSD for simplification. For a jointly proper Gaussian process, the existence of the PSD implies ergodicity [103]. As the channel fading process $\{h_k\}$ is assumed to be stationary, $S_h(f)$ is real-valued. Because of the limitation of the velocity of the transmitter, the receiver, and of objects in the environment, the spread of the PSD is limited, and we assume it to be compactly supported within the

interval $[-f_d, f_d]$, with $0 < f_d < 0.5$, i.e., $S_h(f) = 0$ for $f \notin [-f_d, f_d]$. The parameter f_d corresponds to the normalized maximum Doppler shift and, thus, indicates the dynamics of the channel. To ensure ergodicity, we exclude the case $f_d = 0$. Following the definition given in [24, Sec. XII.2, p. 564], this fading channel is sometimes referred to as *nonregular*.[1]

For technical reasons in some of the proofs, i.e., the calculation of the upper bound on the achievable data rate in Chapter 3, its extensions to the MIMO and the frequency selective case in Chapter 8 and Chapter 9, and the derivation of the lower bound on the achievable rate with joint processing of pilot and data symbols in Chapter 7, we restrict to autocorrelation functions $r_h(l)$ which are absolutely summable, i.e.,

$$\sum_{l=-\infty}^{\infty} |r_h(l)| < \infty \tag{2.4}$$

instead of the more general class of square summable autocorrelation functions, i.e.,

$$\sum_{l=-\infty}^{\infty} |r_h(l)|^2 < \infty \tag{2.5}$$

which is already fulfilled due to our assumption that the PSD exists. However, the assumption of absolutely summable autocorrelation functions is not a severe restriction. E.g., the important rectangular PSD, see (2.8), can be arbitrarily closely approximated by a PSD with the shape corresponding to the transfer function of a raised cosine filter, whose corresponding autocorrelation function is absolutely summable.

2.1 Rayleigh Fading and Jakes' Model

The assumed Rayleigh fading model is a commonly used fading model which reasonably describes the channel observed in mobile urban environments with many scattering objects and no line of sight, see, e.g., [15]. It has to be mentioned that the Rayleigh model only describes small scale fading. Large scale fading due to path loss and shadowing is not described by this model and is also outside the scope of this thesis.

Due to the assumption of relative motion with constant velocity between transmitter, receiver, and objects in the environment, the fading becomes temporally correlated. The normalized continuous-time autocorrelation function is given by [54], [112]

---

[1] For a discussion on the justification if a physical channel fading process is *nonregular* based on *real world numbers* see [66].

$$r(t) = \sigma_h^2 J_0 \left(2\pi f_d \frac{t}{T_{\mathrm{Sym}}} \right) \tag{2.6}$$

where $J_0(\cdot)$ is a zeroth-order Bessel function of the first kind. The corresponding PSD $S_h(f)$ in (2.3) of the discrete-time fading process is given by

$$S_h(f)\big|_{\mathrm{Jakes}} = \begin{cases} \frac{\sigma_h^2}{\pi\sqrt{f_d^2 - f^2}} & \text{for } |f| < f_d \\ 0 & \text{for } f_d \le |f| \le 0.5 \end{cases}. \tag{2.7}$$

These correlation properties can be derived analytically for a dense scatterer environment with a vertical receive antenna with a constant azimuthal gain, a uniform distribution of signals arriving at all angles, i.e., in the interval $[0, 2\pi)$, and with uniformly distributed phases based on a sum of sinusoids [54]. The sum-of-sinusoids model can also be used to generate temporally correlated Rayleigh fading for simulation.

Sometimes the Jakes' PSD in (2.7) is approximated by the following rectangular PSD

$$S_h(f)\big|_{\mathrm{rect}} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \le f_d \\ 0 & \text{for } f_d < |f| \le 0.5 \end{cases}. \tag{2.8}$$

This approximation entails only a small difference with respect to the performance of the corresponding channel estimation [6], [79, pp. 651 and 658]. The performance of the channel estimation can be measured by the estimation error variance, which is related to the capacity. For mathematical tractability, we will also use the rectangular PSD in (2.8) for several derivations.

As already stated, the discrete-time autocorrelation function $r_h(l)$ corresponding to a rectangular PSD $S_h(f)\big|_{\mathrm{rect}}$, which is given by

$$r_h(l) = \sigma_h^2 \mathrm{sinc}(2f_d l) \tag{2.9}$$

is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a shape corresponding to the transfer function of a raised cosine (RC) filter, i.e.,

$$S_h(f)\big|_{\mathrm{RC}} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \le (1 - \beta_{\mathrm{ro}})f_d \\ \frac{\sigma_h^2}{4f_d}\left[1 - \sin\left(\frac{\frac{2\pi}{2f_d}(f - f_d)}{2\beta_{\mathrm{ro}}} \right) \right] & \text{for } (1 - \beta_{\mathrm{ro}})f_d < f \le (1 + \beta_{\mathrm{ro}})f_d \\ 0 & \text{for } f_d(1 + \beta_{\mathrm{ro}}) < |f| \le 0.5 \end{cases}. \tag{2.10}$$

Here $0 \le \beta_{\mathrm{ro}} \le 1$ is the roll-off factor. For $\beta_{\mathrm{ro}} \to 0$ the PSD $S_h(f)\big|_{\mathrm{RC}}$ approaches the rectangular PSD $S_h(f)\big|_{\mathrm{rect}}$. Furthermore, the discrete-time autocorrelation function corresponding to $S_h(f)\big|_{\mathrm{RC}}$ is given by

$$r_h(l) = \sigma_h^2 \text{sinc}(2f_d l) \frac{\cos\left(\beta_{\text{ro}} \pi 2 f_d l\right)}{1 - 4\beta_{\text{ro}}^2 4 f_d^2 l^2} \tag{2.11}$$

which for $\beta_{\text{ro}} > 0$ is absolutely summable. Thus, the rectangular PSD in (2.8) can be arbitrarily closely approximated by a PSD with an absolutely summable autocorrelation function. Therefore, in the rest of this work, we often evaluate the derived bounds on the achievable rate for a rectangular PSD of the channel fading process, although some of the derivations are based on the assumption of an absolutely summable autocorrelation function.

Typical fading channels, as they are observed in mobile communication environments, are characterized by relatively small normalized Doppler frequencies f_d in the regime of $f_d \ll 0.1$. Therefore, the restriction to channels with $f_d < 0.5$, i.e., *nonregular* fading, in the present work is reasonable. Depending on the carrier frequency f_c, the relative velocity between transmitter and receiver v, and the symbol rate $1/T_{\text{Sym}}$, the maximum normalized Doppler frequency is given by

$$f_d = \frac{v \cdot f_c}{c} \cdot T_{\text{Sym}} \tag{2.12}$$

where $c \approx 2.998 \cdot 10^8$ m/s is the speed of light. Considering parameters of a typical mobile communication system, e.g., $f_c = 2$ GHz, a maximum velocity of $v = 300$ km/h, and a symbol rate of 1 MHz leads to a maximum Doppler frequency of only $f_d \approx 0.00056$.

Notice that although in practical scenarios the observed channel dynamics are very small, within this work we always consider the range of $0 < f_d < 0.5$ to get a thorough understanding of the behavior of the bounds on the achievable rate.

2.2 Matrix-Vector Notation

We base the derivation of bounds on the achievable rate on the following matrix-vector notation of the system model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{2.13}$$

where the vectors are defined as

$$\mathbf{y} = [y_1, \ldots, y_N]^T \tag{2.14}$$

$$\mathbf{x} = [x_1, \ldots, x_N]^T \tag{2.15}$$

$$\mathbf{n} = [n_1, \ldots, n_N]^T. \tag{2.16}$$

The matrix $\mathbf{H}$ is diagonal and defined as $\mathbf{H} = \text{diag}(\mathbf{h})$ with $\mathbf{h} = [h_1, \ldots, h_N]^T$. Here the $\text{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal ele-

ments are given by the argument vector. The diagonal matrix $\mathbf{X}$ is given by $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. The quantity N is the number of considered symbols. Later on, we investigate the case of $N \to \infty$ to evaluate the achievable rate.

Using this vector notation, we express the temporal correlation of the fading process by the correlation matrix

$$\mathbf{R}_h = \mathrm{E}[\mathbf{h}\mathbf{h}^H] \tag{2.17}$$

which has a Hermitian Toeplitz structure.

Concerning the input distribution, unless otherwise stated, we make the assumption that the symbols x_k are i.d. , with an maximum average power σ_x^2. For the nominal mean SNR we introduce the variable

$$\rho = \frac{\sigma_x^2 \sigma_h^2}{\sigma_n^2}. \tag{2.18}$$

Notice that we use the term *nominal* mean SNR as ρ only corresponds to the actual mean SNR in case σ_x^2 is the average transmit power. For the case of a non-peak power constrained input distribution, the achievable rate is maximized by using the maximum average transmit power σ_x^2. Thus, in the non-peak power constrained case ρ corresponds to the actual mean SNR.

2.3 Limitations of the Model

In this section, we discuss the limitations of the model. Therefore, let us consider an appropriately bandlimited continuous-time model first, where the channel output is given by

$$y(t) = h(t)s(t) + n(t) \tag{2.19}$$

with $h(t)$ being the continuous-time channel fading process, i.e., the corresponding discrete-time process h_k is given by

$$h_k = h(kT_{\mathrm{Sym}}) \tag{2.20}$$

where T_{Sym} is the symbol duration. Analogously, the continuous-time and the discrete-time additive noise and channel output processes are related by

$$n_k = n(kT_{\mathrm{Sym}}) \tag{2.21}$$

$$y_k = y(kT_{\mathrm{Sym}}). \tag{2.22}$$

The continuous-time transmit process $s(t)$ is given by

$$s(t) = \sum_{k=-\infty}^{\infty} x_k g(t - kT_{\text{Sym}}) \tag{2.23}$$

where $g(t)$ is the transmit pulse. We assume the use of bandlimited transmit pulses, which have an infinite impulse response[2]. In typical systems often root-raised cosine pulses are used such that in combination with the matched filter at the receiver intersymbol interference is minimized. Their impulse response $g(t)$ and their normalized frequency response $G(f)$ are given by

$$G(f) = \sqrt{G_{\text{RC}}(f)} \tag{2.24}$$

with $G_{\text{RC}}(f)$ being the transfer function of the raised cosine filter, cf. (2.10)

$$G_{\text{RC}}(f) = \begin{cases} T_{\text{Sym}} & \text{for } |f| \le \frac{1-\beta_{\text{ro}}}{2} \\ \frac{T_{\text{Sym}}}{2}\left[1 + \cos\left(\frac{\pi}{\beta_{\text{ro}}}\left[|f| - \frac{1-\beta_{\text{ro}}}{2}\right]\right)\right] & \text{for } \frac{1-\beta_{\text{ro}}}{2} < |f| \le \frac{1+\beta_{\text{ro}}}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

and [3]

$$g(t) = \begin{cases} 1 - \beta_{\text{ro}} + 4\frac{\beta_{\text{ro}}}{\pi} & \text{for } t = 0 \\ \frac{\beta_{\text{ro}}}{\sqrt{2}}\left[\left(1 + \frac{2}{\pi}\right)\sin\left(\frac{\pi}{4\beta_{\text{ro}}}\right) + \left(1 - \frac{2}{\pi}\right)\cos\left(\frac{\pi}{4\beta_{\text{ro}}}\right)\right] & \text{for } t = \pm\frac{T_{\text{Sym}}}{4\beta_{\text{ro}}} \\ \frac{\sin\left[\pi\frac{t}{T_{\text{Sym}}}(1-\beta_{\text{ro}})\right] + 4\beta_{\text{ro}}\frac{t}{T_{\text{Sym}}}\cos\left[\pi\frac{t}{T_{\text{Sym}}}(1+\beta_{\text{ro}})\right]}{\pi\frac{t}{T_{\text{Sym}}}\left[1 - \left(4\beta_{\text{ro}}\frac{t}{T_{\text{Sym}}}\right)^2\right]} & \text{otherwise} \end{cases}. \tag{2.26}$$

In the following, we assume a roll-off factor of $\beta_{\text{ro}} = 0$ corresponding to sinc transmit pulses and, thus, we have a rectangular PSD with a normalized bandwidth of 1.

The continuous-time input/output relation in (2.19) has the following stochastic representation in frequency domain

$$S_y(f) = S_h(f) \star S_s(f) + S_n(f) \tag{2.27}$$

where $\star$ denotes convolution and $S_y(f)$, $S_h(f)$, $S_s(f)$, and $S_n(f)$ are the normalized power spectral densities of the continuous-time processes $y(t)$, $h(t)$, $s(t)$, and $n(t)$, e.g.,

$$S_s(f) = \int_{-\infty}^{\infty} \text{E}\left[s(t + \tau)s^*(t)\right] e^{-j2\pi f\tau} d\tau \tag{2.28}$$

---

[2] For a discussion on the contradiction that physical signals must be bandlimited but on the other hand are not infinite in time the interested reader is referred to [113].

and correspondingly for the other PSDs. Here we always assume normalization with $1/T_{\mathrm{Sym}}$.

We are interested in the normalized bandwidth of the component $S_h(f) \star S_s(f)$, i.e., the component containing information on the transmitted sequence $\{x_k\}$. The normalized bandwidth of the transmit signal $s(t)$ directly corresponds to the normalized bandwidth of the transmit pulse $g(t)$, which is assumed to be 1, see above. The normalized bandwidth of the channel fading process is given by $2f_d$. Thus, the normalized bandwidth of the component $S_h(f) \star S_s(f)$ is given by $1 + 2f_d$.

To get a sufficient statistic, we would have to sample the channel output $y(t)$ at least with a frequency of $\frac{1+2f_d}{T_{Sym}}$. As the discrete-time channel output process $\{y_k\}$ is a sampled version of $y(t)$ with the rate $1/T_{\mathrm{Sym}}$, the discrete-time observation process $\{y_k\}$ is not a sufficient statistic of $y(t)$. As usually the normalized maximum Doppler frequency f_d is very small, the amount of discarded information is negligible.

All further derivations are based on the discrete-time model and therefore, are not based on a sufficient statistic, i.e., information is discarded. Beside the fact that in realistic systems the dynamics is very small and, thus, the amount of discarded information is small, in typical systems channel estimation is also performed at symbol rate signals and therefore also exhibits the loss due to the lack of a sufficient statistic. In addition, much of the current literature on the study of the capacity of stationary Rayleigh fading channels, e.g., [67], and on the achievable rate with synchronized detection, e.g., [6], is based on symbol rate discrete-time input-output relations and therefore do not ask the question about a sufficient statistic. However, this should not be understood as a motivation to use the symbol rate signal model. Furthermore, these considerations should be kept in mind in the later evaluations, especially, as we examine the derived bounds not only for very small values of f_d.

2.4 Operational and Information Theoretic Capacity

In this section, we first briefly introduce the definition of information theoretic capacity given by Shannon and recall the channel coding theorem [108] linking information theoretic capacity to operational capacity. Furthermore, we recall results on the extension of Shannon's definition of capacity, which was restricted to memoryless channels, to channels with memory, like the stationary fading channel considered in the present work.

For a memoryless channel the *information theoretic capacity* C_{info} is defined as the maximum mutual information[3] $\mathcal{I}(y;x)$ between the channel input x and the channel output y while maximizing over the distribution of the in-

---

[3] Notice that Shannon [108] defined the capacity without using the mutual information, but directly based on the definition of entropy.

put sample x, i.e., see also [19]

$$C_{\text{info}} = \max_{p(x)} \mathcal{I}(y; x) \tag{2.29}$$

where the mutual information is given by

$$\mathcal{I}(y; x) = h(y) - h(y|x) \tag{2.30}$$

and $h(\cdot)$ is the differential entropy defined by

$$h(y) = \mathrm{E}_y \left[\log \left(p(y)\right)\right] \tag{2.31}$$
$$h(y|x) = \mathrm{E}_{y,x} \left[\log \left(p(y|x)\right)\right]. \tag{2.32}$$

Based on (2.30) the mutual information $\mathcal{I}(y; x)$ can be understood as a measure about the information the random variable y contains about the random variable x, or alternatively on the reduction of the uncertainty on y when knowing x.

Besides this mathematical definition of capacity the main contribution of Shannon was the channel coding theorem. With the channel coding theorem Shannon proved that for memoryless channels all rates below the capacity C_{op}—which in this context often is named *operational capacity*— are achievable, i.e., for each rate $R < C_{\text{op}}$ there exists a code for which the probability of an erroneously decoded codeword approaches zero in the limit of an infinite codeword length. Conversely, this means that error-free transmission is only possible at rates R with $R \leq C_{\text{op}}$. The channel coding theorem states that for memoryless channels, information theoretic and operational capacity coincide. Therefore, the capacity C for a memoryless channel is given by

$$C \equiv C_{\text{info}} = C_{\text{op}}. \tag{2.33}$$

Depending on the specific type of channel, it is necessary to introduce further constraints, to get a finite capacity. E.g., for the additive white Gaussian noise (AWGN) channel the capacity is infinite in the case of an unconstrained input power. Therefore, usually its capacity is given based on a constraint on the maximum average input power P_{av}.

In the context of the stationary flat-fading channel model given in (2.1) and (2.13), we introduce the following constraints on the average power P_{av} and the peak power P_{peak},

$$\frac{1}{N}\mathrm{E}\left[\mathbf{x}^H\mathbf{x}\right] \leq P_{\text{av}} \tag{2.34}$$
$$\max_{1 \leq k \leq N} |x_k|^2 \leq P_{\text{peak}}. \tag{2.35}$$

The average power constraint will be always used in the following derivations, while we will use the peak power constraint only at specific places.

It is important to note that Shannon's theorem is based on the assumption of memoryless channels, i.e., all usages of the channel are independent. For the stationary fading channel considered in the present work, this assumption does not hold, as the channel fading process is temporally correlated. Due to this temporal correlation, e.g., the channel observation y_k also contains information on the channel fading weight h_{k-1} and thus on the previous transmit symbol x_{k-1}.

The coincidence of information theoretic capacity and operational capacity can be extended to channels with memory under some further conditions [103]. Before discussing this, we introduce the definition of information theoretic capacity in the context of the stationary fading channel given in (2.13).

The information theoretic capacity per unit time of the stationary fading channel model is given by

$$C_{\text{info}} = \lim_{N \to \infty} \sup_{\mathcal{P}_{\text{gen}}} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}). \tag{2.36}$$

where the supremum is taken over the set $\mathcal{P}_{\text{gen}}$ of input distributions given by

$$\mathcal{P}_{\text{gen}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, \frac{1}{N} \mathrm{E}[\mathbf{x}^H \mathbf{x}] \leq P_{\text{av}}, \max_{1 \leq k \leq N} |x_k|^2 \leq P_{\text{peak}} \right\}. \tag{2.37}$$

The definition of the information theoretic capacity holds whenever the limit in (2.36) exists.

The peak-power constraint in (2.37) is not generally necessary. Only some of the following derivations are based on a peak power constraint. The case of an unconstrained peak-power corresponds to $P_{\text{peak}} = \infty$ in (2.37).

Corresponding to the memoryless channel, in case of the stationary fading channel the operational capacity C_{op} corresponds to the maximum achievable rate R, which implies the existence of a code with a decoding error probability that approaches zero for infinite codeword length, i.e., $N \to \infty$.

Now we recall the conditions required for the coincidence of information theoretic and operational capacity in case of a channel with memory given in [103]. To describe these conditions, we quote the following definitions on weakly mixing and ergodic processes given in [103], which itself cites [74, Section 5] and [91, p.70].

Define $\phi_i(z_1, z_2, \ldots, z_n)$ with $i = 1, 2$ to be two bounded measurable functions of an arbitrary number of complex variables $z_1, \ldots, z_n$. Furthermore, define the operator M_t as $\lim_{t \to \infty} \frac{1}{t} \sum_1^t$ for a discrete-time process $\{z_k\}$. In addition we define $\psi(t)$ as

$$\psi(t) = \mathrm{E}\left[\phi_1(z_{t_1}, \ldots, z_{t_n}) \cdot \phi_2(z_{t_1^*+t}, \ldots, z_{t_n^*+t})\right] \\ - \mathrm{E}\left[\phi_1(z_{t_1}, \ldots, z_{t_n}) \cdot \phi_2(z_{t_1^*}, \ldots, z_{t_n^*})\right]. \tag{2.38}$$

A stationary stochastic process $\{z_k\}$ is

- weakly mixing if, for all choices of ϕ_1, ϕ_2 and times $t_1, \ldots, t_n, t_1^*, \ldots, t_n^*$

$$M_t \left[\psi^2(t) \right] = 0 \tag{2.39}$$

- ergodic if, for all choices of ϕ_1, ϕ_2 and times $t_1, \ldots, t_n, t_1^*, \ldots, t_n^*$

$$M_t \left[\psi(t) \right] = 0. \tag{2.40}$$

Notice, that an ergodic process is also weakly mixing. Based on concepts concerning information stability and the Shannon-McMillan-Breiman theorem for finite-alphabet ergodic sources, the following proposition is derived in [103]:

Proposition [103]: If the processes $\{h_k\}$ and $\{n_k\}$ are stationary weakly mixing, and if $\{h_k\}$, $\{n_k\}$, and $\{x_k\}$ are mutually independent, then for every $P_{\mathrm{av}}, P_{\mathrm{peak}} > 0$ the information theoretic capacity C_{info} is well defined and corresponds to the operational capacity C_{op}.

As we assume that the PSD of the channel fading process $S_h(f)$ in (2.3) exists, and as the fading process is assumed to be jointly proper Gaussian, the channel fading process is ergodic. For a discussion on this relation see [103]. For proper Gaussian processes, ergodicity is equivalent to weakly mixing. Thus, for the system model (2.13) considered in this work, operational and information theoretic capacity coincide. This allows us to use the term of *information theoretic capacity* in the following.

2.4.1 Outage Capacity

For completeness of presentation, we also mention that there exist further capacity measures. The preceding definition of information theoretic capacity (*Shannon capacity*) considers the maximum rate being achievable with a probability of an decoding error that approaches zero for infinitely long codewords. If we deviate from the focus on an arbitrary small probability of error, we can also use the definition of outage capacity.

The $q\%$-outage capacity C_{outage} is defined as the information rate that is guaranteed for $(100 - q)\%$ of the channel realizations [84], i.e.,

$$P(C \leq C_{\mathrm{outage}}) = q\%. \tag{2.41}$$

Therefore, this definition is especially interesting in the context of channels, where the channel quality changes over time like fading channels.

However, within the rest of this work we will not use the measure *outage capacity* but will restrict to the use of the *information theoretic capacity*.

Chapter 3
Bounds on the Achievable Rate of a Flat-Fading Channel

Based on the model of a discrete-time stationary Rayleigh flat-fading channel introduced in the previous chapter, we calculate bounds on the achievable rate with identically distributed (i.d.) input symbols within the present chapter. This assumption will be needed in the derivation of the upper bound on the achievable rate. Concerning the input distribution, we explicitly discuss the following two cases: On the one hand, the case of a peak-power constraint and, on the other hand, the special case of i.i.d. zero-mean proper Gaussian input symbols. At the end of this chapter, we also compare the bounds on the achievable rate with i.d. input symbols to existing bounds on the channel capacity given in [67] and [107]. Parts of the present chapter have already been published in [28].

Based on the coincidence of operational and information theoretic capacity as discussed in Section 2.4 and the ergodicity of the channel fading process, the capacity of the stationary fading channel with an average power constraint is given by

$$C = \lim_{N \to \infty} \sup_{\mathcal{P}} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}) = \sup_{\mathcal{P}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \tag{3.1}$$

where $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ is the mutual information rate between the channel input and the channel output defined as

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \lim_{N \to \infty} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}). \tag{3.2}$$

Furthermore, the set $\mathcal{P}$ contains all probability density functions $p(\mathbf{x})$ over the complex space $\mathbb{C}^N$ with an average power $P_{\mathrm{av}} \leq \sigma_x^2$, i.e.,

$$\mathcal{P} = \left\{ p(\mathbf{x}) \big| \mathbf{x} \in \mathbb{C}^N, \frac{1}{N} \mathrm{E}[\mathbf{x}^H \mathbf{x}] \leq \sigma_x^2 \right\}. \tag{3.3}$$

Note that for the RHS of (3.1) we make a slight misuse of notation. The set $\mathcal{P}$ is defined for input vectors $\mathbf{x}$ of length N. Therefore, the exchange of the limit and the supremum in (3.1), as it is used when expressing C based on the mutual information rate, is formally not correct. However, to avoid a further complication of notation, we use the set $\mathcal{P}$ also in the context of information rates. The same holds also in the following for other sets of input distributions.

As already described in Chapter 2, if not otherwise stated, we assume i.d. input symbols. As this assumption is not proven to be optimal, in the following we use the term *achievable rate*, which we define as

$$R = \sup_{\mathcal{P}_{\text{i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \qquad (3.4)$$

where $\mathcal{P}_{\text{i.d.}}$ is the set over all probability density functions (PDFs) $p(\mathbf{x})$ with i.d. elements, i.e.,

$$\mathcal{P}_{\text{i.d.}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, \quad p(x_i) = p(x_j) \quad \forall i, j, \quad \mathrm{E}[|x_k|^2] \le \sigma_x^2 \quad \forall k \right\}. \quad (3.5)$$

At some specific points, we will further constrain the input distribution to be peak power limited. In this case, the set of input distributions is given by

$$\mathcal{P}_{\text{i.d.}}^{\text{peak}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, \; p(x_i) = p(x_j) \quad \forall i, j, \right.$$

$$\left. \left\{ \mathrm{E}[|x_k|^2] \le \sigma_x^2, \; p(x_k) = 0 \; \forall x_k : |x_k|^2 > P_{\text{peak}} \right\} \forall k \right\} \quad (3.6)$$

which corresponds to the set $\mathcal{P}_{\text{i.d.}}$ in (3.5) with a further restriction on the peak power, i.e., $|x_k|^2 \le P_{\text{peak}}$.

3.1 The Mutual Information Rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$

In general, by means of the chain rule, the mutual information rate in (3.2) can be expanded as [8]

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y}) \qquad (3.7)$$

where $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is the mutual information rate in case the channel is known at the receiver, i.e., the mutual information rate of the coherent channel, and $\mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y})$ is the penalty due to the channel uncertainty. It is interesting to note that the penalty term can be further separated as follows

$$\mathcal{I}'(\mathbf{x};\mathbf{h}|\mathbf{y}) \overset{(a)}{=} \mathcal{I}'(\mathbf{y},\mathbf{x};\mathbf{h}) - \mathcal{I}'(\mathbf{y};\mathbf{h})$$

$$\overset{(b)}{=} \mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x}) + \mathcal{I}'(\mathbf{h};\mathbf{x}) - \mathcal{I}'(\mathbf{y};\mathbf{h})$$

$$\overset{(c)}{=} \mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x}) - \mathcal{I}'(\mathbf{y};\mathbf{h}) \tag{3.8}$$

where for (a) and (b) we use the chain rule for mutual information and for (c) we exploit the fact that the mutual information between the channel fading process described by $\mathbf{h}$ and the input sequence $\mathbf{x}$ is zero due to the independency of $\mathbf{h}$ and $\mathbf{x}$ and, thus,

$$\mathcal{I}'(\mathbf{h};\mathbf{x}) = 0. \tag{3.9}$$

Obviously, with (3.8) the penalty term corresponds to the difference between the knowledge on the channel $\mathbf{h}$ that can be obtained from the observation $\mathbf{y}$ while knowing the transmit sequences $\mathbf{x}$ in comparison to not knowing it.

Within this chapter, we derive bounds on the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ based on the following straightforward separation into differential entropy rates,

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}). \tag{3.10}$$

Here, $h'(\cdot)$ indicates the differential entropy rate and is defined as

$$h'(\cdot) = \lim_{N\to\infty} \frac{1}{N} h(\cdot). \tag{3.11}$$

In Section 3.2, we give a lower and an upper bound on the channel output entropy rate $h'(\mathbf{y})$, which are independent of the PSD of the channel fading process $S_h(f)$. In contrast, in Section 3.3 we derive an upper bound and a lower bound on $h'(\mathbf{y}|\mathbf{x})$, where the upper bound holds for an arbitrary PSD of the channel fading process with compact support. For the lower bound on $h'(\mathbf{y}|\mathbf{x})$ we find a closed form expression only for the special case of a rectangular PSD. For these derivations, we only assume i.d. input symbols. Further restrictions on the input distribution, like a peak power constraint or the assumption on zero-mean proper Gaussian input symbols are applied later when evaluating the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ based on the bounds on $h'(\mathbf{y})$ and $h'(\mathbf{y}|\mathbf{x})$ in Section 3.4.

3.2 The Received Signal Entropy Rate $h'(\mathbf{y})$

In this section, we derive an upper bound and a lower bound on $h'(\mathbf{y})$, which are independent of the PSD of the channel fading process $S_h(f)$.

3.2.1 Lower Bound on $h'(\mathbf{y})$

The mutual information with perfect channel state information at the receiver can be upper-bounded by

$$\mathcal{I}(\mathbf{y}; \mathbf{x}|\mathbf{h}) = h(\mathbf{y}|\mathbf{h}) - h(\mathbf{y}|\mathbf{h}, \mathbf{x})$$
$$\leq h(\mathbf{y}) - h(\mathbf{y}|\mathbf{h}, \mathbf{x}). \tag{3.12}$$

Here we make use of the fact that conditioning reduces entropy. Thus, we can lower-bound the entropy rate $h'(\mathbf{y})$ by

$$h'(\mathbf{y}) \geq \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}). \tag{3.13}$$

The mutual information rate in case the channel is known at the receiver, i.e., the first term on the RHS of (3.13), depends obviously on the input distribution. Thus, we postpone its further evaluation to Section 3.4.1.3 and Section 3.4.2 where we derive the bounds on the achievable rate.

The second term on the RHS of (3.13) originates from AWGN and, thus, can be calculated as

$$h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) = \log\left(\pi e \sigma_n^2\right). \tag{3.14}$$

Hence, a lower bound on the entropy rate $h'(\mathbf{y})$ is given by

$$h'(\mathbf{y}) \geq h'_L(\mathbf{y}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + \log\left(\pi e \sigma_n^2\right). \tag{3.15}$$

3.2.2 Upper Bound on $h'(\mathbf{y})$

In this section, we give an upper bound on the entropy rate $h'(\mathbf{y})$. First, we make use of the fact that the entropy $h(\mathbf{y})$ of a zero-mean complex random vector $\mathbf{y}$ of dimension N with nonsingular correlation matrix $\mathbf{R}_y = \mathrm{E}[\mathbf{y}\mathbf{y}^H]$ is upper-bounded by [82]

$$h(\mathbf{y}) \leq \log\left[(\pi e)^N \det(\mathbf{R}_y)\right]. \tag{3.16}$$

By Hadamard's inequality, the determinant of a matrix is upper-bounded by the product of its diagonal entries. The diagonal entries of $\mathbf{R}_y$ are given by $\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2$, where $\alpha\sigma_x^2$ is the average transmit power with $\alpha \in [0, 1]$. Thus, if $\alpha < 1$ an average power smaller than the maximum average power σ_x^2 allowed by the set of input distributions $\mathcal{P}_{\text{i.d.}}$ is used by the transmitter.[1]

Therefore, further upper-bounding the RHS of (3.16) yields

---

[1] Notice that in case we would assume i.i.d. transmit symbols, the upper-bounding of the RHS of (3.16) by Hadamard's inequality is not required, as in this case the correlation matrix $\mathbf{R}_y$ is already diagonal with $\mathbf{R}_y = (\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2)\mathbf{I}_N$.

$$h(\mathbf{y}) \leq N \log \left[\pi e (\alpha \sigma_x^2 \sigma_h^2 + \sigma_n^2) \right]. \tag{3.17}$$

Hence, with (3.17) the upper bound $h'_U(\mathbf{y})$ on the entropy rate $h'(\mathbf{y})$ is given by

$$h'(\mathbf{y}) \leq h'_U(\mathbf{y}) = \log \left(\pi e \left(\alpha \sigma_x^2 \sigma_h^2 + \sigma_n^2 \right) \right). \tag{3.18}$$

The discussion of the parameter α is postponed until the study of the mutual information rate.

In Appendix A.1, we give another upper bound on $h'(\mathbf{y})$ for the specific case of zero-mean proper Gaussian inputs based on numerical integration to calculate $h(y_k)$, i.e., the output entropy at an individual time instant, based on the output distribution $p(y_k)$, see [88]. As this bound can only be evaluated numerically using Hermite polynomials and Simpson's rule or by Monte Carlo integration, we do not further consider it here.

3.3 The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$

In this section, we give an upper bound and a lower bound on the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$. We recall the derivation of the upper bound given in [21]. The upper bound holds for an arbitrary PSD of the channel fading process with compact support. In contrast, to the best of our knowledge, the lower bound on $h'(\mathbf{y}|\mathbf{x})$ is new and only holds for the special case of a rectangular PSD.

3.3.1 Upper Bound on $h'(\mathbf{y}|\mathbf{x})$

The probability density of $\mathbf{y}$ conditioned on $\mathbf{x}$ is zero-mean proper Gaussian. Therefore, its entropy is

$$h(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{\mathbf{x}} \left[\log \left((\pi e)^N \det(\mathbf{R}_{y|x}) \right) \right] \tag{3.19}$$

where the covariance matrix $\mathbf{R}_{y|x}$ is given by

$$\mathbf{R}_{y|x} = \mathrm{E}_{\mathbf{h},\mathbf{n}} \left[\mathbf{y}\mathbf{y}^H | \mathbf{x} \right] = \mathrm{E}_{\mathbf{h}} \left[\mathbf{X}\mathbf{h}\mathbf{h}^H \mathbf{X}^H | \mathbf{x} \right] + \sigma_n^2 \mathbf{I}_N$$
$$= \mathbf{X}\mathbf{R}_h \mathbf{X}^H + \sigma_n^2 \mathbf{I}_N. \tag{3.20}$$

As the channel correlation matrix $\mathbf{R}_h$ is Hermitian and thus normal, the spectral decomposition theorem applies, i.e.,

$$\mathbf{R}_h = \mathbf{U}\boldsymbol{\Lambda}_h \mathbf{U}^H \tag{3.21}$$

where the diagonal matrix $\mathbf{\Lambda}_h = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_N)$ contains the eigenvalues λ_i of $\mathbf{R}_h$ and the matrix $\mathbf{U}$ is unitary.

The following upper-bounding of $h(\mathbf{y}|\mathbf{x})$ is already known from [21]. Making use of (3.21), Jensen's inequality and the concavity of the log function, we can upper-bound $h(\mathbf{y}|\mathbf{x})$ in (3.19) as follows

$$h(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X}\mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H \mathbf{X}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \quad (3.22)$$

$$\overset{(a)}{=} \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X}\mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \quad (3.23)$$

$$\overset{(b)}{\leq} \log \det \left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) + N \log(\pi e \sigma_n^2)$$

$$= \log \det \left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \mathbf{\Lambda}_h + \mathbf{I}_N \right) + N \log(\pi e \sigma_n^2)$$

$$= \sum_{i=1}^{N} \log \left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \lambda_i + 1 \right) + N \log(\pi e \sigma_n^2) \quad (3.24)$$

where $\alpha \sigma_x^2$ is the average transmit power and, thus, $\alpha \in [0, 1]$. For (a) the following relation is used

$$\det(\mathbf{A}\mathbf{B} + \mathbf{I}) = \det(\mathbf{B}\mathbf{A} + \mathbf{I}) \quad (3.25)$$

which holds as $\mathbf{A}\mathbf{B}$ has the same eigenvalues as $\mathbf{B}\mathbf{A}$ for $\mathbf{A}$ and $\mathbf{B}$ being square matrices [52, Theorem 1.3.20]. For (b) we have used the fact that $\log \det(\cdot)$ is concave on the set of positive definite matrices[2].

To calculate the bound on the entropy rate $h'(\mathbf{y}|\mathbf{x})$, we consider the case $N \to \infty$, i.e., the dimension of the matrix $\mathbf{\Lambda}_h$ grows without bound. As $\mathbf{R}_h$ is Hermitian Toeplitz, we can evaluate (3.24) using Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [43], [44]. Consequently,

---

[2] For the special case of independent transmit symbols, (b) can also be shown in two steps by using Jensen's inequality and in a second step expressing the determinant by a Laplacian expansion by minors to calculate the expectation, i.e.,

$$\mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X}\mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right] \leq \log \mathrm{E}_{\mathbf{x}} \left[\det \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X}\mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right]$$

$$= \log \det \left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \mathbf{U}\mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \quad (3.26)$$

The assumption on i.i.d. input symbols can be made as the upper bound on $h(\mathbf{y}|\mathbf{x})$ is used to establish a lower bound on capacity, which is given by the mutual information rate for an arbitrary input distribution. Furthermore, the supremum of the lower bound on $h'(\mathbf{y})$ in (3.15) over all input distributions fulfilling the specific power constraints, which is required for the calculation of the lower bound on the achievable rate, is also achieved for i.i.d. input symbols, see Section 3.4.2.

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \lambda_i + 1\right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(S_h(f)\frac{\alpha \sigma_x^2}{\sigma_n^2} + 1\right) df. \qquad (3.27)$$

Notice that due to the assumption that the PSD (2.3) exists, the condition

$$\sum_{m=-\infty}^{\infty} |r_h(m)|^2 < \infty \qquad (3.28)$$

holds and, thus, Szegö's theorem can be applied.

Hence, we get the following upper bound

$$h'(\mathbf{y}|\mathbf{x}) \leq h'_U(\mathbf{y}|\mathbf{x})$$
$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(S_h(f)\frac{\alpha \sigma_x^2}{\sigma_n^2} + 1\right) df + \log(\pi e \sigma_n^2). \qquad (3.29)$$

The discussion of the factor α, i.e., the average transmit power $\alpha \sigma_x^2$, is postponed until studying the mutual information rate.

At this point, it is interesting to note that for constant modulus (CM) input symbols the differential entropy rate $h'(\mathbf{y}|\mathbf{x})$ is equal to the upper bound $h'_U(\mathbf{y}|\mathbf{x})$, i.e.,

$$h'(\mathbf{y}|\mathbf{x})\big|_{\text{CM}} = h'_U(\mathbf{y}|\mathbf{x}) \qquad (3.30)$$

as in this case (3.23) simplifies due to the following relation

$$\mathbf{X}^H\mathbf{X}\big|_{\text{CM}} = \alpha \sigma_x^2 \mathbf{I} \qquad (3.31)$$

with the transmit power given by $\alpha \sigma_x^2$ and, thus, (b) succeeding (3.23) holds with equality.

3.3.2 Lower Bound on $h'(\mathbf{y}|\mathbf{x})$ for a Rectangular PSD

In this section, we give a lower bound on the entropy rate $h'(\mathbf{y}|\mathbf{x})$ for the special case of a rectangular PSD, which is a common approximation of the actual PSD in typical system design.

For the further proof, we derive a circulant matrix which is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h$. Hereby, we follow a specific approach as shown in [43, Section 4.4], where the circulant matrix is constructed by sampling the PSD of the channel fading process. For the discussion of the asymptotic equivalence, we write $\mathbf{R}_h^{(N)}$ instead of $\mathbf{R}_h$, where the superscript (N) denotes the size of the square matrix $\mathbf{R}_h$.

Let the first column of the circulant matrix $\mathbf{C}_h^{(N)}$ be given by

$$\left(c_0^{(N)} \; c_1^{(N)} \; \cdots \; c_{N-1}^{(N)} \right)^T \tag{3.32}$$

where again the superscript (N) denotes the size of the square matrix $\mathbf{C}_h^{(N)}$. The elements $c_k^{(N)}$ are given by

$$c_k^{(N)} = \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h \left(\frac{l}{N} \right) e^{j2\pi k \frac{l}{N}} \tag{3.33}$$

where $\tilde{S}_h(f)$ is the periodic continuation of $S_h(f)$ given in (2.3), i.e.,

$$\tilde{S}_h(f) = \sum_{k=-\infty}^{\infty} \delta(f-k) \star S_h(f) \tag{3.34}$$

and $S_h(f)$ being zero outside the interval $|f| \leq 0.5$ for which it is defined in (2.3).

As we assume that the autocorrelation function of the channel fading process is absolutely summable, the PSD of the channel fading process $S_h(f)$ is Riemann integrable, and it holds that

$$
\begin{aligned}
\lim_{N \to \infty} c_k^{(N)} &= \lim_{N \to \infty} \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h \left(\frac{l}{N} \right) e^{j2\pi k \frac{l}{N}} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f) e^{j2\pi k f} \, df \\
&= r_h(k)
\end{aligned}
\tag{3.35}
$$

with $r_h(k)$ given by (2.2).

As the eigenvectors of a circulant matrix are given by a discrete Fourier transform (DFT), the eigenvalues $\tilde{\lambda}_m^{(N)}$ with $m = 1, \ldots, N$ of the circulant matrix $\mathbf{C}_h^{(N)}$ are given by

$$\tilde{\lambda}_m^{(N)} = \sum_{k=0}^{N-1} c_k^{(N)} e^{-j2\pi k \frac{m-1}{N}} \tag{3.36}$$

and, thus,

$$c_k^{(N)} = \frac{1}{N} \sum_{m=1}^{N} \tilde{\lambda}_m^{(N)} e^{j2\pi k \frac{m-1}{N}} \tag{3.37}$$

so that with (3.33) we get

$$\tilde{\lambda}_m^{(N)} = \tilde{S}_h\left(\frac{m-1}{N}\right). \tag{3.38}$$

Consequently, the spectral decomposition of the circulant matrix $\mathbf{C}_h^{(N)}$ is given by

$$\mathbf{C}_h^{(N)} = \mathbf{F}^{(N)}\tilde{\mathbf{\Lambda}}_h^{(N)}\left(\mathbf{F}^{(N)}\right)^H \tag{3.39}$$

where the matrix $\mathbf{F}^{(N)}$ is a unitary DFT matrix, i.e., its elements are given by

$$\left[\mathbf{F}^{(N)}\right]_{k,l} = \frac{1}{\sqrt{N}}e^{j2\pi\frac{(k-1)(l-1)}{N}}. \tag{3.40}$$

Furthermore, the matrix $\tilde{\mathbf{\Lambda}}_h^{(N)}$ is diagonal with the elements $\tilde{\lambda}_m^{(N)}$ given in (3.38).

It can be shown that the Toeplitz matrix $\mathbf{R}_h^{(N)}$ and the circulant matrix $\mathbf{C}_h^{(N)}$ are asymptotically equivalent in case of an absolutely summable autocorrelation function $r_h(l)$, see [43, Lemma 4.6]. In the context of proving this lemma it is shown that the weak norm of the difference of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ converges to zero as $N \to \infty$, i.e.,

$$\lim_{N\to\infty}\left|\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)}\right| = 0 \tag{3.41}$$

where the weak norm of a matrix $\mathbf{B}$ is defined as

$$|\mathbf{B}| = \left(\frac{1}{N}\text{Tr}\left[\mathbf{B}^H\mathbf{B}\right]\right)^{\frac{1}{2}}. \tag{3.42}$$

The convergence of the weak norm of the difference $\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)}$ towards zero is required later on.

By the construction of the circulant matrix $\mathbf{C}_h^{(N)}$, the eigenvalues $\tilde{\lambda}_m^{(N)}$ of $\mathbf{C}_h^{(N)}$ are given by (3.38), i.e.,

$$\tilde{\lambda}_m^{(N)} = \begin{cases} S_h\left(f = \frac{m-1}{N}\right) & \text{for } 1 \le m \le \lceil\frac{N}{2}\rceil \\ S_h\left(f = \frac{m-1}{N} - 1\right) & \text{for } \lceil\frac{N}{2}\rceil < m \le N \end{cases}. \tag{3.43}$$

Thus, if the PSD of the channel fading process $S_h(f)$ is rectangular, the eigenvalues of the circulant matrix $\mathbf{C}_h^{(N)}$ are given by[3]

---

[3] Recall that a rectangular PSD $S_h(f)$ corresponds to $r_h(l) = \sigma_h^2\text{sinc}(2f_d l)$ which is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a raised cosine shape, whose corresponding correlation function is absolutely summable, see Section 2.1 for a discussion.

$$\tilde{\lambda}_m^{(N)} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } 1 \leq m \leq f_d N + 1 \vee (1 - f_d)N + 1 \leq m \leq N \\ 0 & \text{otherwise} \end{cases} . \tag{3.44}$$

Now, we want to apply the asymptotic equivalence of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ to lower-bound the entropy rate $h'(\mathbf{y}|\mathbf{x})$ given by

$$h'(\mathbf{y}|\mathbf{x}) = \lim_{N \to \infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x}) \tag{3.45}$$

with $h(\mathbf{y}|\mathbf{x})$ given in (3.22). Thus, we have to show that

$$\lim_{N \to \infty} \frac{1}{N} \mathbf{E_x} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X} \mathbf{R}_h^{(N)} \mathbf{X}^H + \mathbf{I}_N \right) \right]$$
$$= \lim_{N \to \infty} \frac{1}{N} \mathbf{E_x} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X} \mathbf{C}_h^{(N)} \mathbf{X}^H + \mathbf{I}_N \right) \right] . \tag{3.46}$$

Using (3.25), equation (3.46) can be rewritten as

$$\lim_{N \to \infty} \frac{1}{N} \mathbf{E_x} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{R}_h^{(N)} + \mathbf{I}_N \right) \right]$$
$$= \lim_{N \to \infty} \frac{1}{N} \mathbf{E_x} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \right) \right] . \tag{3.47}$$

To prove (3.47) we have to show that the matrices

$$\mathbf{K}_1^{(N)} = \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{R}_h^{(N)} + \mathbf{I}_N \tag{3.48}$$

$$\mathbf{K}_2^{(N)} = \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \tag{3.49}$$

are asymptotically equivalent [43, Theorem 2.4]. This means that we have to show that both matrices are bounded in the strong norm, and that the weak norm of their difference converges to zero for $N \to \infty$ [43, Section 2.3].

Concerning the condition with respect to the strong norm we have to show that

$$\left\| \mathbf{K}_1^{(N)} \right\| < \infty \tag{3.50}$$

$$\left\| \mathbf{K}_2^{(N)} \right\| < \infty \tag{3.51}$$

with the strong norm of the matrix $\mathbf{B}$ defined by

$$\|\mathbf{B}\|^2 = \max_k \gamma_k \tag{3.52}$$

where γ_k are the eigenvalues of the Hermitian nonnegative definite matrix $\mathbf{B}\mathbf{B}^H$. The diagonal matrix $\mathbf{X}^H\mathbf{X}$ contains the transmit powers of the indi-

vidual transmit symbols. In the case of Gaussian input distributions, for a given $\epsilon > 0$, there exists a finite value $M(\epsilon)$ such that the transmit power is smaller than $M(\epsilon)$ with probability $1 - \epsilon$. In addition, the strong norms of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ are bounded, too. Concerning the boundedness of the eigenvalues of the Hermitian Toeplitz matrix $\mathbf{R}_h^{(N)}$ see [43, Lemma 4.1]. Thus, the strong norms of $\mathbf{K}_1^{(N)}$ and $\mathbf{K}_2^{(N)}$ are asymptotically almost surely bounded, i.e., with probability converging to 1.

Furthermore, for the weak norm of the difference $\mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)}$ we get for $N \to \infty$

$$
\begin{aligned}
\left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &= \left| \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{R}_h^{(N)} + \mathbf{I}_N - \left(\frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \right) \right| \\
&= \left| \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \left(\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right) \right| \\
&\overset{(a)}{\leq} \frac{1}{\sigma_n^2} \left\| \mathbf{X}^H \mathbf{X} \right\| \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right|
\end{aligned}
\tag{3.53}
$$

where for (a) we have used [43, Lemma 2.3]. Based on the above argumentation that $\left\| \mathbf{X}^H \mathbf{X} \right\|$ is bounded with probability one, we get for $N \to \infty$

$$
\begin{aligned}
\lim_{N \to \infty} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &\leq \lim_{N \to \infty} \frac{1}{\sigma_n^2} \left\| \mathbf{X}^H \mathbf{X} \right\| \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \\
&= 0
\end{aligned}
\tag{3.54}
$$

due to (3.41). Thus, we have proved that (3.46) holds and we can express the entropy rate $h'(\mathbf{y}|\mathbf{x})$ by

$$
\begin{aligned}
h'(\mathbf{y}|\mathbf{x}) &= \lim_{N \to \infty} \frac{1}{N} \left[\mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X} \mathbf{C}_h^{(N)} \mathbf{X}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \right] \\
&= \lim_{N \to \infty} \frac{1}{N} \left[\mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X} \mathbf{F} \tilde{\mathbf{\Lambda}}_h \mathbf{F}^H \mathbf{X}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \right].
\end{aligned}
\tag{3.55}
$$

Here $\mathbf{F} \tilde{\mathbf{\Lambda}}_h \mathbf{F}^H$ is the spectral decomposition of the circulant matrix $\mathbf{C}_h$ (from here on we omit the superscript (N) for ease of notation), see (3.39). Thus, $\tilde{\mathbf{\Lambda}}_h$ is a diagonal matrix containing the eigenvalues $\tilde{\lambda}_k$ as given in (3.44) and the matrix $\mathbf{F}$ is a unitary matrix with the eigenvectors of $\mathbf{C}_h$ on its columns.

To calculate a lower bound on $h'(\mathbf{y}|\mathbf{x})$ we transform the term in the brackets at the RHS of (3.55) as follows

$$\mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{X} \mathbf{F} \tilde{\mathbf{\Lambda}}_h \mathbf{F}^H \mathbf{X}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2)$$

$$\overset{(a)}{=} \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \tilde{\mathbf{\Lambda}}_h \mathbf{F}^H \mathbf{X}^H \mathbf{X} \mathbf{F} + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2)$$

$$\overset{(b)}{=} \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{\sigma_h^2}{2 f_d \sigma_n^2} \tilde{\mathbf{F}}^H \mathbf{X}^H \mathbf{X} \tilde{\mathbf{F}} + \mathbf{I}_{2\lfloor f_d N \rfloor + 1} \right) \right] + N \log(\pi e \sigma_n^2) \quad (3.56)$$

where for (a) we have used (3.25). For (b), the eigenvalue distribution in (3.44) is used, and the matrix $\tilde{\mathbf{F}}$ is given by

$$\tilde{\mathbf{F}} = \left[\mathbf{f}_1, \ldots, \mathbf{f}_{\lfloor f_d N + 1 \rfloor}, \mathbf{f}_{\lceil (1 - f_d) N + 1 \rceil}, \ldots, \mathbf{f}_N \right] \in \mathbb{C}^{N \times (2\lfloor f_d N \rfloor + 1)} \quad (3.57)$$

where the $\mathbf{f}_i$ are the orthonormal columns of the unitary matrix $\mathbf{F}$. Now, we apply the following inequality given in [46, Lemma 1].

Lemma 3.1. *Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ with orthonormal rows and $m \leq n$. Then*

$$\log \det \left(\mathbf{A} \operatorname{diag}(p_1, \ldots, p_n) \mathbf{A}^H \right) \geq \operatorname{trace} \left[\mathbf{A} \operatorname{diag}(\log p_1, \ldots, \log p_n) \mathbf{A}^H \right] \quad (3.58)$$

if $p_1, \ldots, p_n > 0$.

With Lemma 3.1, we can lower-bound (3.56) such that

$$\mathrm{E}_{\mathbf{x}} \left[\log \det \left(\frac{\sigma_h^2}{2 f_d \sigma_n^2} \tilde{\mathbf{F}}^H \mathbf{X}^H \mathbf{X} \tilde{\mathbf{F}} + \mathbf{I}_{2\lfloor f_d N \rfloor + 1} \right) \right] + N \log(\pi e \sigma_n^2)$$

$$\geq \mathrm{E}_{\mathbf{x}} \left[\operatorname{trace} \left[\tilde{\mathbf{F}}^H \operatorname{diag} \left(\log \left(\frac{\sigma_h^2 |x_1|^2}{2 f_d \sigma_n^2} + 1 \right), \ldots, \log \left(\frac{\sigma_h^2 |x_N|^2}{2 f_d \sigma_n^2} + 1 \right) \right) \tilde{\mathbf{F}} \right] \right]$$
$$+ N \log(\pi e \sigma_n^2)$$

$$= \operatorname{trace} \left[\tilde{\mathbf{F}}^H \operatorname{diag} \left(\mathrm{E}_x \log \left(\frac{\sigma_h^2 |x_1|^2}{2 f_d \sigma_n^2} + 1 \right), \ldots, \mathrm{E}_x \log \left(\frac{\sigma_h^2 |x_N|^2}{2 f_d \sigma_n^2} + 1 \right) \right) \tilde{\mathbf{F}} \right]$$
$$+ N \log(\pi e \sigma_n^2)$$

$$\overset{(a)}{=} \sum_{k=1}^{2\lfloor f_d N \rfloor + 1} \mathrm{E}_x \log \left(\frac{\sigma_h^2}{2 f_d \sigma_n^2} |x|^2 + 1 \right) + N \log(\pi e \sigma_n^2) \quad (3.59)$$

where (a) results, because all x_k are identically distributed and because the columns of $\tilde{\mathbf{F}}$ are orthonormal.

Hence, with (3.55) the entropy rate is given by

$$h'(\mathbf{y}|\mathbf{x}) = \lim_{N\to\infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x})$$

$$\geq \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{2\lfloor f_d N\rfloor + 1} \mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) + \log(\pi e\sigma_n^2)$$

$$= 2f_d\mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) + \log(\pi e\sigma_n^2) = h'_L(\mathbf{y}|\mathbf{x}). \qquad (3.60)$$

Thus, we have found a general lower bound on the entropy rate $h'(\mathbf{y}|\mathbf{x})$ for i.d. input distributions. To the best of our knowledge, this is the only known lower bound on the entropy rate $h'(\mathbf{y}|\mathbf{x})$ which is not based on a peak power constraint. Therefore, this bound later on enables the calculation of an upper bound on the achievable rate with i.d. input symbols with zero-mean proper Gaussian distribution.

3.3.2.1 Discussion on the Assumption of a Rectangular PSD

For the case of constant modulus (CM) input distributions, it can be shown that the rectangular PSD maximizes $h'(\mathbf{y}|\mathbf{x})$ among all PSDs with a compact support interval $[-f_d, f_d]$ and a channel power σ_h^2. Therefor, we have to calculate $\sup_{S_h(f)\in\mathcal{S}} h'(\mathbf{y}|\mathbf{x})|_{\mathrm{CM}}$ where the set $\mathcal{S}$ of PSDs is given by

$$\mathcal{S} = \left\{ S_h(f) = 0 \text{ for } f_d < |f| \leq 0.5, \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f)df = \sigma_h^2 \right\}. \qquad (3.61)$$

With (3.29) and (3.30) we get

$$\sup_{S_h(f)\in\mathcal{S}} h'(\mathbf{y}|\mathbf{x})|_{\mathrm{CM}} = \sup_{S_h(f)\in\mathcal{S}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(S_h(f)\frac{\alpha\sigma_x^2}{\sigma_n^2} + 1\right) df + \log(\pi e\sigma_n^2)$$

$$= \sup_{S_h(f)\in\mathcal{S}} \int_{-f_d}^{f_d} \log\left(S_h(f)\frac{\alpha\sigma_x^2}{\sigma_n^2} + 1\right) df + \log(\pi e\sigma_n^2)$$
$$(3.62)$$

$$\overset{(a)}{=} \int_{-f_d}^{f_d} \log\left(\frac{\sigma_h^2}{2f_d}\frac{\alpha\sigma_x^2}{\sigma_n^2} + 1\right) df + \log(\pi e\sigma_n^2) \qquad (3.63)$$

i.e., the PSD $S_h(f)$ which maximizes $h'(\mathbf{y}|\mathbf{x})$ is rectangular

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{otherwise} \end{cases}. \qquad (3.64)$$

The last step in (3.63) can be proven as follows. The $S_h(f)$ that maximizes (3.62) has to maximize the following functional J for all f and constant c

$$J(S_h) = \int_{-f_d}^{f_d} \log\left(S_h(f)\frac{\alpha\sigma_x^2}{\sigma_n^2} + 1\right) df + c\left(\int_{-f_d}^{f_d} S_h(f)df - \sigma_h^2\right) \quad (3.65)$$

where the last term accounts for the constraint

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f)df = \sigma_h^2. \quad (3.66)$$

Therefore, the following equation must be fulfilled for each f within the interval $[-f_d, f_d]$

$$\frac{\partial J}{\partial S_h(f)} = \frac{\frac{\alpha\sigma_x^2}{\sigma_n^2}}{S_h(f)\frac{\alpha\sigma_x^2}{\sigma_n^2} + 1} + c = 0. \quad (3.67)$$

As this equation has to be fulfilled for each f and constant c, $S_h(f)$ must be constant within the interval $[-f_d, f_d]$. Thus, with (3.66), (3.63) follows.

We conjecture that a rectangular PSD of the channel fading process maximizes $h'(\mathbf{y}|\mathbf{x})$ for any i.d. input distribution with an average power $\alpha\sigma_x^2$. Concerning this discussion see also [11, Section IV-A]. Consequently, the lower bound in (3.60) then holds only for a rectangular PSD. As this lower bound on $h'(\mathbf{y}|\mathbf{x})$ is finally used for the upper bound on $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ and following the preceding conjecture, we get an upper bound on the achievable rate for a given maximum Doppler spread f_d for the worst case PSD.

3.4 The Achievable Rate

Based on the upper and lower bounds on $h'(\mathbf{y})$ and $h'(\mathbf{y}|\mathbf{x})$, we are now able to give upper and lower bounds on the achievable rate with i.d. inputs.

3.4.1 Upper Bound

3.4.1.1 Peak Power Constrained Input

First, we consider input distributions with a peak power constraint, i.e., $|x_k|^2 \leq P_{\text{peak}}$. For this case, we can upper-bound the supremum of $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ over the set of input distribution $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$ in (3.6) as follows

$$\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \{h_U'(\mathbf{y}) - h_L'(\mathbf{y}|\mathbf{x})\} \quad (3.68)$$

with $h_U'(\mathbf{y})$ and $h_L'(\mathbf{y}|\mathbf{x})$ given by (3.18) and (3.60).

The calculation of the supremum in (3.68) is done in two steps. The inner supremum is taken over the constrained set $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha$ being characterized by an average power $\alpha\sigma_x^2$ which holds with equality. Because of the fact that in (3.6) we only use a constraint on the maximum average input power given by σ_x^2 the outer supremum is taken over $\alpha \in [0,1]$. The set $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha$ is given by

$$
\mathcal{P}_{\text{i.d.}}^{\text{peak}}\Big|\alpha = \left\{ p(\mathbf{x}) \Big| \mathbf{x} \in \mathbb{C}^N, \;\; p(x_i) = p(x_j) \quad \forall i,j, \right.
$$

$$
\left. \left\{ \mathrm{E}[|x_k|^2] = \alpha\sigma_x^2, \;\; p(x_k) = 0 \; \forall x_k : |x_k|^2 > P_{\text{peak}} \right\} \forall k \right\}
$$
(3.69)

which corresponds to the set $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$ except that the average power is now fixed to $\alpha\sigma_x^2$ with equality. Such a separation has also been used in [107] and in [33].

Based on the split of the supremum into two parts (3.68) becomes

$$
\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \left\{ h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \right\}
$$

$$
= \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \left\{ \log\left(\alpha\rho + 1\right) - 2f_d \mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1 \right) \right\}
$$

$$
= \sup_{\alpha \in [0,1]} \left\{ \log\left(\alpha\rho + 1\right) - 2f_d \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1 \right) \right\}
$$
(3.70)

with the nominal average SNR ρ given in (2.18).

The term containing the infimum on the RHS of (3.70) can be lower-bounded in the following way

$$
\inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1 \right)
$$

$$
= \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \int_{|x|=0}^{\sqrt{P_{\text{peak}}}} \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1 \right)}{|x|^2} |x|^2 p(|x|)d|x|
$$

$$
\overset{(a)}{=} \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}P_{\text{peak}} + 1 \right)}{P_{\text{peak}}} \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha} \int_{|x|=0}^{\sqrt{P_{\text{peak}}}} |x|^2 p(|x|)d|x|
$$

$$= \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}P_{\text{peak}} + 1\right)}{P_{\text{peak}}}\alpha\sigma_x^2 \tag{3.71}$$

where for (a) we have used that all factors of the integrand are positive and that the term

$$\frac{1}{|x|^2}\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) = \frac{1}{z}\log\left(cz + 1\right) \tag{3.72}$$

with $c = \frac{\sigma_h^2}{2f_d\sigma_n^2}$ and $z = |x|^2$ is monotonically decreasing in z as

$$\frac{\partial}{\partial z}\left\{\frac{1}{z}\log\left(cz + 1\right)\right\} = \frac{c}{(cz+1)z} - \frac{\log(cz+1)}{z^2} \leq 0$$

$$\Leftrightarrow \frac{cz}{cz+1} \leq \log\left(cz + 1\right) \tag{3.73}$$

which holds for $cz > -1$. Thus, the term in (3.72) is minimized for $z = |x|^2 = P_{\text{peak}}$. A similar approach to calculate the infimum in (3.71) has been used in [33] and [127] for analogous problems. Notice that the result given in (3.71) means that the infimum on $h_L'(\mathbf{y}|\mathbf{x})$ for a fixed average transmit power is achieved with on-off keying.

With (3.71), we get the following upper bound on the RHS of (3.70)

$$\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \sup_{\alpha\in[0,1]}\left\{\log\left(\alpha\rho + 1\right) - 2f_d\frac{\alpha\sigma_x^2}{P_{\text{peak}}}\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}P_{\text{peak}} + 1\right)\right\}$$

$$= \sup_{\alpha\in[0,1]}\left\{\log\left(\alpha\rho + 1\right) - 2f_d\frac{\alpha}{\beta}\log\left(\frac{1}{2f_d}\rho\beta + 1\right)\right\} \tag{3.74}$$

with the *nominal peak-to-average power ratio*[4]

$$\beta = \frac{P_{\text{peak}}}{\sigma_x^2}. \tag{3.75}$$

As the argument of the supremum on the RHS of (3.74) is concave in α and, thus, there exists a unique maximum, it can easily be shown that the supremum in (3.74) with respect to $\alpha \in [0, 1]$ is given by

$$\alpha_{\text{opt}} = \min\left\{1, \left(\frac{2f_d}{\beta}\log\left(\frac{1}{2f_d}\rho\beta + 1\right)\right)^{-1} - \frac{1}{\rho}\right\} \tag{3.76}$$

---

[4] Instead of the common term *peak-to-average power ratio* we choose the term *nominal peak-to-average power ratio*, as in case of a peak power constraint, it is not necessarily optimal to use the maximum average power σ_x^2. In case the actual average power is equal to the maximum average power σ_x^2, β corresponds to the actual peak-to-average power ratio.

and, thus,

$$\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \log\left(\alpha_{\text{opt}}\rho + 1\right) - 2f_d \frac{\alpha_{\text{opt}}}{\beta} \log\left(\frac{1}{2f_d}\rho\beta + 1\right) \tag{3.77}$$

$$= \mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{P_{\text{peak}}}. \tag{3.78}$$

Hence, we have found an upper bound on the achievable rate with i.d. input symbols and a peak power constraint for the special case of a rectangular PSD of the channel fading process. Note that the writing $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{P_{\text{peak}}}$ denotes an upper bound on the peak power constrained achievable rate. We will use similar notations in the following.

Furthermore, notice that the case $\alpha_{\text{opt}} < 1$ corresponds to the case that it is not optimal to use the maximum average transmit power allowed by the set $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$. This behavior is a result of the peak power constraint. Therefore, consider the extreme case $\beta = 1$ and $f_d = 0.5$, i.e., an uncorrelated channel. $\alpha = 1$ then would correspond to constant modulus signaling, i.e., the transmitter puts all information into the phase of the transmitted signal. As the channel is uncorrelated from symbol to symbol and unknown to the receiver, the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ is zero. Therefore, it is better, if the receiver does not use all its transmit power, i.e., uses an $\alpha < 1$, enabling modulation of the magnitude, which leads to a positive $\mathcal{I}'(\mathbf{y}; \mathbf{x})$.

The choice $\alpha_{\text{opt}} = 1$, corresponding to the case that it is optimal to use the maximum possible average transmit power, can be shown to be optimal, on the one hand, if

$$1 \leq \rho \leq \frac{2f_d}{\beta}\left[\exp\left(\frac{1}{2}\frac{\beta}{2f_d}\right) - 1\right] \tag{3.79}$$

or, on the other hand, if

$$2f_d \leq \frac{\beta}{\rho + 2} \text{ for } \rho \leq 1. \tag{3.80}$$

For a proof of these conditions see Appendix A.2. As in realistic scenarios f_d is close to zero, the condition (3.79) and (3.80) are typically fulfilled. However, for the parameter range displayed in Fig. 3.1 the conditions in (3.79) and (3.80) are not always fulfilled.

The upper bound in (3.78) is a special case of the upper bound on the peak power constrained capacity given in [107]. For a further comparison of both bounds we refer to Section 3.6.

For a non-peak power constrained upper bound on the achievable rate with i.d. inputs we have to evaluate the following expression

$$\sup_{\mathcal{P}_{\text{i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \le \sup_{\mathcal{P}_{\text{i.d.}}} \left\{ h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \right\}$$

$$\le \sup_{\alpha \in [0,1]} \left\{ \log\left(\alpha\rho + 1\right) - 2f_d \inf_{\mathcal{P}_{\text{i.d.}}|\alpha} \mathrm{E}_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) \right\} \tag{3.81}$$

where $\mathcal{P}_{\text{i.d.}}|\alpha$ corresponds to $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|\alpha$ in (3.69) but without having a peak power constraint. From the calculation of the infimum in (3.71) it is obvious that in the non-peak power constrained case the infimum on the RHS of (3.81) becomes zero and, thus,

$$\sup_{\mathcal{P}_{\text{i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \le \log\left(\rho + 1\right) \tag{3.82}$$

which obviously is loose as it is the capacity of an AWGN channel being already larger than the coherent capacity of the fading channel.

3.4.1.2 Gaussian Input Distribution

Considering higher order modulation, the nominal peak-to-average power ratio β may become relatively large. For proper Gaussian inputs it is in fact infinite. Obviously, for large peak powers P_{peak}, the second term on the RHS of (3.77) approaches zero and it can be shown that this bound becomes loose for $\beta > 1$ and high SNR, see Section 3.4.4. Therefore, we now follow a different approach. As in the case of perfect receiver side channel knowledge, i.e., the coherent scenario, an i.i.d. zero-mean proper Gaussian input distribution is capacity-achieving, we guess that it will not be highly non-optimal for low channel dynamics. Thus, we now calculate an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. As for the derivation of the lower bound $h'_L(\mathbf{y}|\mathbf{x})$ in (3.60) no assumptions on a peak power constraint have been made, it still holds for probability distribution functions with an unbounded support, and we are able to evaluate it for the case of zero-mean proper Gaussian input symbols. The lower bound on the entropy rate $h'_L(\mathbf{y}|\mathbf{x})$ in (3.60) is for zero-mean proper Gaussian transmit symbols with an average transmit power of $\alpha\sigma_x^2$ given by

$$h'_L(\mathbf{y}|\mathbf{x})\big|_{\text{PG},\alpha\sigma_x^2} = 2f_d \int_0^\infty \log\left(\frac{\sigma_h^2\alpha\sigma_x^2}{2f_d\sigma_n^2}z + 1\right) e^{-z}dz + \log(\pi e \sigma_n^2). \tag{3.83}$$

Thus, for i.i.d. zero-mean proper Gaussian (PG) input symbols, the achievable rate is upper bounded by

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} = \sup_{\alpha \in [0,1]} \left\{ \log\left(\alpha\rho + 1\right) - 2f_d \int_0^\infty \log\left(\frac{\alpha\rho}{2f_d}z + 1\right) e^{-z} dz \right\}$$

$$\overset{(a)}{=} \log\left(\rho + 1\right) - 2f_d \int_0^\infty \log\left(\frac{\rho}{2f_d}z + 1\right) e^{-z} dz. \tag{3.84}$$

where (a) follows by showing that the argument of the supremum monotonically increases with α, as

$$\frac{\partial}{\partial\alpha} \left\{ \log\left(\alpha\rho + 1\right) - 2f_d \int_0^\infty \log\left(\frac{\alpha\rho}{2f_d}z + 1\right) e^{-z} dz \right\}$$

$$= \frac{\rho}{\alpha\rho + 1} - 2f_d \int_0^\infty \frac{\frac{\rho}{2f_d}z}{\frac{\alpha\rho}{2f_d}z + 1} e^{-z} dz$$

$$\overset{(b)}{\geq} \frac{\rho}{\alpha\rho + 1} - 2f_d \frac{\frac{\rho}{2f_d}}{\frac{\alpha\rho}{2f_d} + 1} \geq 0. \tag{3.85}$$

For (b) we use that $\frac{\frac{\rho}{2f_d}z}{\frac{\alpha\rho}{2f_d}z+1}$ is concave in z and, thus, we can apply Jensen's inequality.

Equation (3.84) indicates that due to the lack of a peak power constraint it is optimal to use the maximum average transmit power σ_x^2, i.e., to choose $\alpha = 1$. Notice that for the derivation of this upper bound the assumption on independent input symbols has not been used. Nevertheless, in the context of proper Gaussian input distributions we will use the term *i.i.d.* as independent input symbols are capacity achieving in the coherent scenario.

To the best of our knowledge, the upper bound in (3.84) is new. Most other available upper bounds on the capacity hold only for input distributions with a peak power constraint and become loose for high peak-to-average power ratios, see, e.g., [105] and [107], and also the upper bound on the achievable rate with i.d. inputs in (3.78), which is equivalent to the capacity upper bound in [107, Proposition 2.2]. However, it has to be stated that the peak power constrained upper bounds in [105] and [107] are upper bounds on capacity and hold for an arbitrary PSD of the channel fading process.

3.4.1.3 Modification based on Coherent Upper Bound

Furthermore, we know that the mutual information rate in case of perfect channel state information at the receiver $\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h})$ always upper-bounds the mutual information rate in the absence of channel state information, i.e.,

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h}). \tag{3.86}$$

Therefore, we can modify the upper bounds as follows

$$\mathcal{I}'_{U_{mod}}(\mathbf{y};\mathbf{x}) = \min\{\mathcal{I}'_U(\mathbf{y};\mathbf{x}), \sup_{\mathcal{P}}\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h})\}. \qquad (3.87)$$

The term $\sup_{\mathcal{P}}\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h})$ corresponds to the coherent capacity in case the average power is constrained to σ_x^2 without any peak power constraint. It is given by

$$
\begin{aligned}
\sup_{\mathcal{P}}\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h}) &= \lim_{N\to\infty} \sup_{\mathcal{P}} \frac{1}{N}\mathrm{E}_{\mathbf{h}}\left[\mathrm{E}_{\mathbf{y},\mathbf{x}}\left[\log\left(\frac{p(\mathbf{y}|\mathbf{h},\mathbf{x})}{p(\mathbf{y}|\mathbf{h})}\right)\middle|\mathbf{h}\right]\right] \\
&\overset{(a)}{=} \sup_{\mathcal{P}} \mathrm{E}_{h_k}\left[\mathrm{E}_{y_k,x_k}\left[\log\left(\frac{p(y_k|h_k,x_k)}{p(y_k|h_k)}\right)\middle|h_k\right]\right] \\
&\overset{(b)}{=} \sup_{\mathcal{P}} \mathcal{I}(y;x|h) \qquad\qquad\qquad (3.88) \\
&\overset{(c)}{=} \mathrm{E}_h\left[\log\left(1+\rho\frac{|h|^2}{\sigma_h^2}\right)\right] \\
&= \int_{z=0}^{\infty} \log\left(1+\rho z\right)e^{-z}dz. \qquad\qquad (3.89)
\end{aligned}
$$

where for (a) we have used that due to conditioning on the channel fading vector $\mathbf{h}$ the channel uses become independent and the supremum is achieved for independent input symbols. Effectively, the coherent Rayleigh flat-fading channel can be interpreted as an AWGN channel with varying SNR. Furthermore, for (b) we dropped the time index for ease of notation. The capacity-achieving input distribution is i.i.d. zero-mean proper Gaussian and the coherent capacity corresponds to an average of the AWGN channel capacity over all fading realizations, which has been used for (c). Obviously, the coherent mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ is independent of the temporal correlation of the channel.

In Fig. 3.1, the upper bound on the achievable rate with a peak power constraint in (3.78) is shown for different nominal peak-to-average power ratios β in comparison to the upper bound on the achievable rate for zero-mean proper Gaussian input symbols in (3.84) (both combined with (3.87)). This comparison shows that except for β close to 1 and a small to average SNR or a sufficiently small channel dynamics the upper bound on the achievable rate for proper Gaussian inputs is lower than the bound based on a peak power constraint in (3.78). In Section 3.4.4 we will show that the peak power constrained upper bound on the achievable rate in (3.78) becomes loose with increasing β.
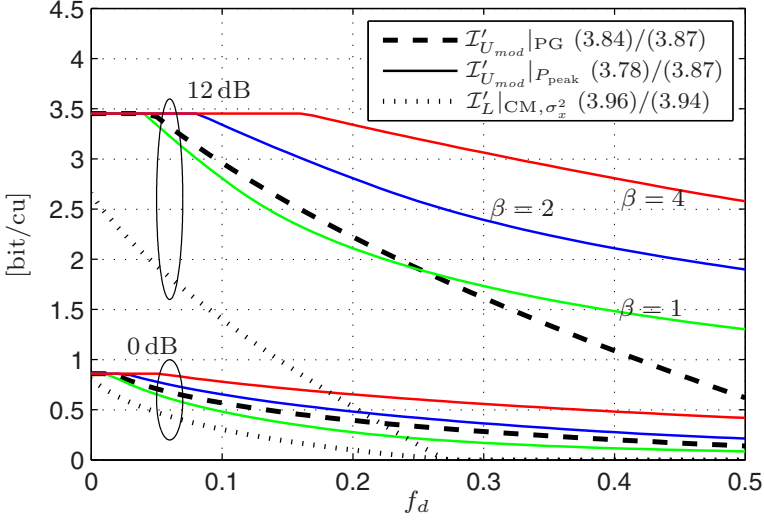
Fig. 3.1: **Comparison of the upper bounds on the achievable rate** for i.d. input symbols with a peak power constraint in (3.78)/(3.87) and with i.i.d. zero-mean proper Gaussian inputs in (3.84)/(3.87) in bits per channel use (cu); (Note, (3.84) also holds for i.d. zero-mean proper Gaussian input symbols); in addition the **peak power constrained lower bound** in (3.96) is shown

3.4.2 Lower Bound

3.4.2.1 No Peak Power Constraint

Using (3.10), (3.15), and (3.29) a lower bound on the capacity is given by

$$\sup_{\mathcal{P}_{\text{i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \geq \sup_{\mathcal{P}_{\text{i.d.}}} \{h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x})\}$$

$$= \sup_{\mathcal{P}_{\text{i.d.}}} \left\{ \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\alpha\sigma_x^2}{\sigma_n^2} S_h(f) + 1\right) df \right\}. \quad (3.90)$$

Notice that lower bounds on the achievable rate are also lower bounds on the capacity. Therefore, in the context of these lower bounds we use the term *capacity* in the following.

It is well known that in case of a non-peak power constrained input distribution the coherent mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is maximized for i.i.d. zero-mean proper Gaussian input symbols. Thus, for an average power of $\alpha\sigma_x^2$ the first term on the RHS of (3.90) is given by, cf. (3.89)

$$\sup_{\mathcal{P}_{\text{i.d.}}|\alpha} \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) = \int_0^\infty \log\left(1 + \alpha\rho z\right) e^{-z} dz = \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})\big|_{\text{PG},\alpha} \qquad (3.91)$$

where $\mathcal{P}_{\text{i.d.}}|\alpha$ corresponds to $\mathcal{P}_{\text{i.d.}}$ in (3.5), but with an average transmit power of $\alpha\sigma_x^2$.

Inserting (3.91) into (3.90) leads to

$$\sup_{\mathcal{P}_{\text{i.d.}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \geq \sup_{\alpha \in [0,1]} \left\{ \int_0^\infty \log\left(\alpha\rho z + 1\right) e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\alpha \frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1\right) df \right\}$$

$$\overset{(a)}{\geq} \int_0^\infty \log\left(\rho z + 1\right) e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1\right) df$$

$$\overset{(b)}{=} \mathcal{I}_L'(\mathbf{y};\mathbf{x})\big|_{\text{PG}} \qquad (3.92)$$

where for (a) we choose $\alpha = 1$, i.e., use the maximum average transmit power, which does not maximize the argument of the supremum in general. Furthermore, note that this lower bound also holds for i.i.d. zero-mean proper Gaussian inputs and, thus, (b) holds. For the special case that $S_h(f)$ is rectangular the lower bound becomes

$$\mathcal{I}_L'(\mathbf{y};\mathbf{x})\big|_{\text{PG}} = \int_0^\infty \log\left(\rho z + 1\right) e^{-z} dz - 2f_d \log\left(\frac{\rho}{2f_d} + 1\right). \qquad (3.93)$$

The bound in (3.92) is already known from [21].

As the mutual information rate is nonnegative, we can further modify the lower bound as follows:

$$\mathcal{I}_{L_{mod}}'(\mathbf{y};\mathbf{x}) = \max\{\mathcal{I}_L'(\mathbf{y};\mathbf{x}), 0\}. \qquad (3.94)$$

3.4.2.2 Peak Power Constraint

Obviously, the lower bound on the achievable rate given in (3.93) does not hold in case of a peak power constrained input, as in this case the coherent mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ is smaller than the RHS of (3.91).

Therefore, in case of a peak power constrained input, we have to evaluate the following expression, see (3.10), (3.14), and (3.15)

$$\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}_L'(\mathbf{y};\mathbf{x}) = \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \left\{ \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) - h_U'(\mathbf{y}|\mathbf{x}) \right\}$$

$$\overset{(a)}{=} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \left\{ \mathcal{I}(y;x|h) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) - h_U'(\mathbf{y}|\mathbf{x}) \right\}. \qquad (3.95)$$

where $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$ is given by (3.6). Furthermore, for (a) we have used that the coherent mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is maximized for independent input symbols, cf. (3.88).

The second term on the RHS of (3.95) $h'(\mathbf{y}|\mathbf{x}, \mathbf{h})$ is independent of the input distribution. Furthermore, it has been shown that for constant modulus input distributions the upper bound $h'_U(\mathbf{y}|\mathbf{x})$ corresponds to the actual entropy rate $h'(\mathbf{y}|\mathbf{x})$, see (3.30).

Recall that the achievable rate for an arbitrary input distribution gives a lower bound on capacity. Thus, based on constant modulus (CM) inputs with a power σ_x^2, and using (3.14), (3.29), and (3.30) a lower bound on the peak power constrained capacity is obviously given by

$$\sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) \geq \mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\text{CM},\sigma_x^2}$$

$$= \mathcal{I}(y; x|h)\big|_{\text{CM},\sigma_x^2} - \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1\right) df \quad (3.96)$$

where $\mathcal{I}(y; x|h)\big|_{\text{CM},\sigma_x^2}$ corresponds to the coherent mutual information using i.i.d. circularly symmetric constant modulus input symbols with power σ_x^2. Hence, we have found a lower bound on the capacity that is achievable with i.d. input symbols. However, as far as we know there is no closed form solution for the first term in (3.96), i.e., $\mathcal{I}(y; x|h)\big|_{\text{CM},\sigma_x^2}$, so it has to be calculated numerically. In addition, for nominal peak-to-average power ratios $\beta > 1$ this bound is in general not tight. The lower bound (3.96) is shown in Fig. 3.1. Obviously, this bound becomes loose with an increasing SNR, as it is based on constant modulus signaling.

3.4.3 Tightness of Bounds on the Achievable Rate

General statements on the tightness of the given bounds are difficult to obtain. In the following we therefore consider on the one hand the case of asymptotically small channel dynamics, and on the other hand we evaluate the tightness of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in general.

3.4.3.1 Asymptotically Small Channel Dynamics

It can be shown that for $f_d \to 0$, the lower bound

$$\mathcal{I}'_L(\mathbf{y}; \mathbf{x}) = h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x})$$
$$= \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x}, \mathbf{h}) - h'_U(\mathbf{y}|\mathbf{x}) \quad (3.97)$$

see (3.13), is equivalent to the mutual information rate in case of perfect channel knowledge

$$\lim_{f_d \to 0} \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) = \mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h}). \tag{3.98}$$

This corresponds to the physical interpretation that a channel that changes arbitrarily slowly can be estimated arbitrarily well, and, therefore, the penalty term $\mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y})$ in (3.7) approaches zero. Thus, for $f_d \to 0$, the lower bound $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$ is tight.

However, notice that this in general not means that $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$ is a tight lower bound on capacity for $f_d \to 0$, as we here do not specify the input distribution. This can be easily observed when studying the peak power constrained lower bound $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\mathrm{CM}, \sigma_x^2}$ in (3.96) which is based on constant modulus signaling. This bound does obviously not converge to the channel capacity for $f_d \to 0$. In contrast, the non-peak power constrained lower bound $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\mathrm{PG}}$ in (3.93) becomes tight for $f_d \to 0$ in the sense that it converges to the coherent capacity. This is also one advantage of our study of bounds on the achievable rate with non-peak power constrained input symbols. As the coherent capacity is achieved by non-peak power constrained input symbols, this approach allows to use a lower bound, which becomes tight for asymptotically small channel dynamics.

3.4.3.2 Gaussian Inputs

Fig. 3.2 shows the upper bound (3.84)/(3.87) and the lower bound (3.93)/(3.94) on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols as a function of the channel dynamics, which is characterized by f_d, for different SNRs. Obviously, the achievable rate strongly decreases with an increasing channel dynamics f_d. Furthermore, the gap between the upper and the lower bound depends on the SNR and gets larger with an increasing SNR. In the following we study the tightness of the given bounds analytically. This examination will show that the gap between the upper and the lower bound is bounded.

To evaluate the tightness of the upper and the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, we first evaluate the tightness of the upper and the lower bound on the channel output entropy rate $h'(\mathbf{y})$ for the special case of proper Gaussian inputs. Afterwards, we evaluate the tightness of the upper and lower bound on $h'(\mathbf{y}|\mathbf{x})$.

The difference between upper bound $h'_U(\mathbf{y})$ and $h'_L(\mathbf{y})$ for i.i.d. zero-mean proper Gaussian inputs with an average power σ_x^2, i.e., $\alpha = 1$, is given by, see (3.15), (3.18), and (3.91)
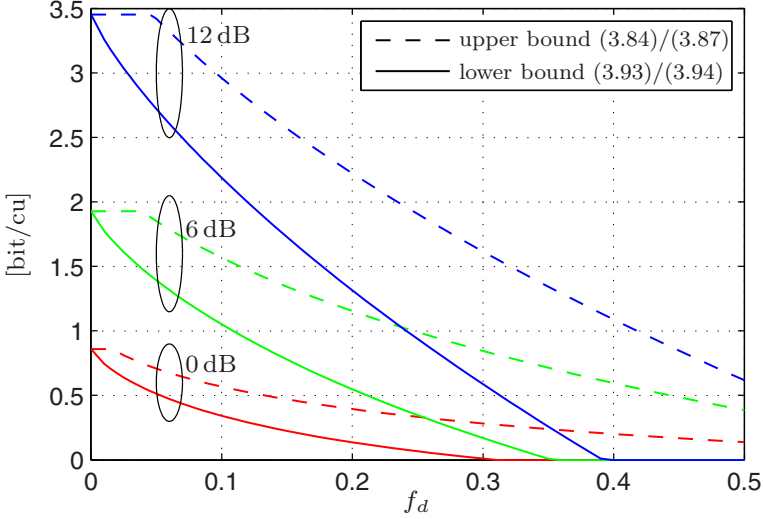
Fig. 3.2: Upper bound and lower bound on the mutual information rate of a Rayleigh flat-fading channel with a rectangular PSD over f_d under the assumption of an i.i.d. zero-mean proper Gaussian input distribution

$$\Delta_{h'(\mathbf{y})} = h'_U(\mathbf{y})\big|_{\alpha=1} - h'_L(\mathbf{y})\big|_{\text{PG},\alpha=1}$$
$$= \log\left(\rho + 1\right) - \int_0^\infty \log\left(1 + \rho z\right) e^{-z} dz. \tag{3.99}$$

Fig. 3.3 shows this difference.

For $\rho \to 0$ the difference $\Delta_{h'(\mathbf{y})}$ converges to zero. For $\rho \to \infty$ the difference is given by

$$\lim_{\rho \to \infty} \Delta_{h'(\mathbf{y})} = \gamma \approx 0.57721 \quad \text{[nat/cu]} \tag{3.100}$$

where γ is the Euler constant. The limit in (3.100) can be found in [70].

The difference $\Delta_{h'(\mathbf{y})}$ monotonically increases with the SNR, which can be shown analogously to (3.85), or alternatively with the proof in Appendix A.3. Thus, $\Delta_{h'(\mathbf{y})}$ is bounded by

$$0 \le \Delta_{h'(\mathbf{y})} \le \gamma. \tag{3.101}$$

The difference between the upper bound and the lower bound on $h'(\mathbf{y}|\mathbf{x})$ in case of a rectangular PSD and the assumption of i.i.d. zero-mean proper Gaussian input symbols with average power σ_x^2 is given by, see (3.29) and (3.83)
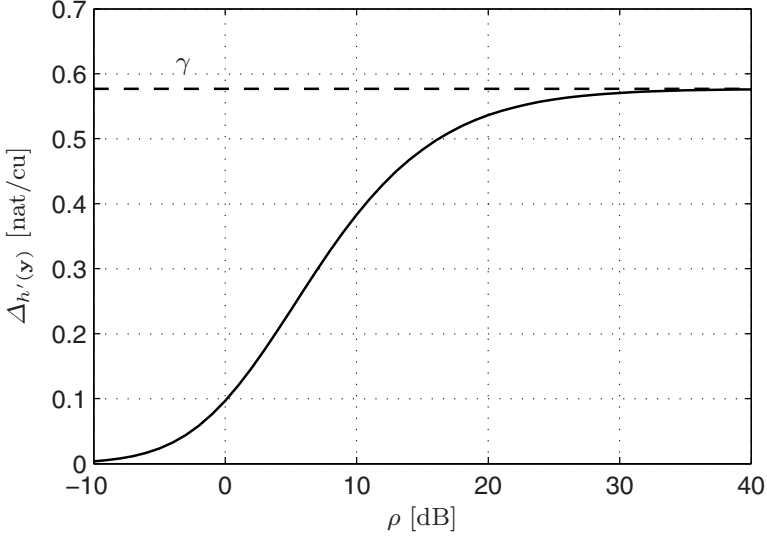
Fig. 3.3: Difference $\Delta_{h'(\mathbf{y})}$ between upper and lower bound on $h'(\mathbf{y})$ for i.i.d. zero-mean proper Gaussian inputs

$$\Delta_{h'(\mathbf{y}|\mathbf{x})} = h'_U(\mathbf{y}|\mathbf{x})\big|_{\alpha=1} - h'_L(\mathbf{y}|\mathbf{x})\big|_{\mathrm{PG},\alpha=1}$$
$$= 2f_d \left\{ \log\left(\frac{\rho}{2f_d} + 1 \right) - \int_0^\infty \log\left(\frac{\rho}{2f_d}z + 1 \right) e^{-z}dz \right\}. \quad (3.102)$$

For asymptotically small Doppler frequencies $\Delta_{h'(\mathbf{y}|\mathbf{x})}$ approaches zero independent of the SNR.

By observing the structural similarity between (3.102) and (3.99), it can be shown that

$$\lim_{\rho \to 0} \Delta_{h'(\mathbf{y}|\mathbf{x})} = 0 \quad (3.103)$$

independent of f_d. For asymptotically high SNR, and a fixed f_d the difference is bounded by

$$\lim_{\rho \to \infty} \Delta_{h'(\mathbf{y}|\mathbf{x})} = 2f_d\gamma \approx 2f_d \cdot 0.57721 \quad [\mathrm{nat/cu}] \quad (3.104)$$

where the same limit as in (3.100) is used.

Corresponding to $\Delta_{h'(\mathbf{y})}$, $\Delta_{h'(\mathbf{y}|\mathbf{x})}$ is monotonically increasing with the SNR and thus, it can be bounded by

$$0 \le \Delta_{h'(\mathbf{y}|\mathbf{x})} \le \gamma 2f_d \quad [\mathrm{nat/cu}]. \quad (3.105)$$

Based on $\Delta_{h'(\mathbf{y})}$ and $\Delta_{h'(\mathbf{y}|\mathbf{x})}$ the difference between the upper bound $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{\mathrm{PG}}$ in (3.84) and the lower bound $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})|_{\mathrm{PG}}$ in (3.93) for i.i.d. zero-mean proper Gaussian input symbols and a rectangular PSD is given by

$$\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} = \mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{\mathrm{PG}} - \mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\mathrm{PG}}$$
$$= \Delta_{h'(\mathbf{y})} + \Delta_{h'(\mathbf{y}|\mathbf{x})}. \tag{3.106}$$

As

$$\lim_{\rho \to 0} \Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} = 0$$
$$\lim_{\rho \to \infty} \Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} = \gamma(1 + 2f_d) \tag{3.107}$$

and as $\Delta_{h'(\mathbf{y})}$, $\Delta_{h'(\mathbf{y})}$, and, thus, $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})}$ monotonically increase with the SNR, we can bound the difference by

$$0 \le \Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} \le \gamma(1 + 2f_d) \quad [\mathrm{nat/cu}]. \tag{3.108}$$

3.4.4 The Asymptotic High SNR Behavior

In this section, we examine the slope of the achievable rate over the SNR for asymptotically large SNRs depending on the channel dynamics[5]. This examination also shows that the peak power constrained upper bound is loose for high SNR and a nominal peak-to-average power ratio $\beta > 1$.

It can be shown that for a compactly supported PSD as defined in Chapter 2 the non-peak power constrained lower bound in (3.92) shows the following behavior

$$\lim_{\rho \to \infty} \frac{\partial \mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\mathrm{PG}}}{\partial \log(\rho)}$$
$$= \lim_{\rho \to \infty} \frac{\partial}{\partial \log(\rho)} \left[\int_{z=0}^{\infty} \log(\rho z + 1) e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_h(f)}{\sigma_h^2} \rho + 1 \right) df \right]$$
$$= \lim_{\rho \to \infty} \left[\int_{z=0}^{\infty} \frac{\rho z}{\rho z + 1} e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\frac{S_h(f)}{\sigma_h^2} \rho}{\frac{S_h(f)}{\sigma_h^2} \rho + 1} df \right]$$
$$= 1 - 2f_d \tag{3.109}$$

as $S_h(f) \ne 0$ for $|f| \le f_d$.

---

[5] When using the term high or large SNR *slope* we refer to the high SNR limit of the derivative of the achievable rate (bound) with respect to the logarithm of the SNR. This quantity is often named *pre-log*.

In [67] it has been shown that the high SNR slope (*pre-log*) of the peak power constrained capacity also corresponds to $1 - 2f_d$, see Section 3.5.

Now, we discuss the high SNR behavior of the upper bound on the achievable rate. We start with the peak power constrained case and restrict here to the special case of using the maximum average power, i.e., $\alpha = 1$, although this is in general not an upper bound on the achievable rate. The motivation for this will become obvious afterwards. For the peak power constrained upper bound given in (3.78) and for the special case $\alpha = 1$ the derivative with respect to $\log(\rho)$ in the high SNR limit is given by

$$
\begin{aligned}
\lim_{\rho \to \infty} \frac{\partial \mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{P_{\text{peak}}, \alpha=1}}{\partial \log(\rho)} &= \lim_{\rho \to \infty} \frac{\partial}{\partial \log(\rho)} \left[\log(\rho + 1) - 2f_d \frac{1}{\beta} \log \left(\frac{\rho \beta}{2f_d} + 1 \right) \right] \\
&= \lim_{\rho \to \infty} \left[\frac{\rho}{\rho + 1} - \frac{2f_d}{\beta} \frac{\frac{\beta}{2f_d} \rho}{\frac{\beta}{2f_d} \rho + 1} \right] \\
&= 1 - \frac{2f_d}{\beta}
\end{aligned}
\tag{3.110}
$$

where β is the nominal peak-to-average power ratio as defined in (3.75). Obviously, if the nominal peak-to-average power ratio β is not equal to one the slope of the peak power constrained upper bound with the constraint $\alpha = 1$ is higher than the slope of the non-peak power constrained lower bound calculated in (3.109), although a further constraint on the input, i.e., the peak power, is introduced. As $\alpha = 1$ does in general not lead to a supremum on $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})$ over the set $\mathcal{P}^{\text{peak}}_{\text{i.d.}}$, (3.110) establishes a lower bound on the high SNR slope of $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{P_{\text{peak}}}$. In combination with the knowledge that the asymptotic high SNR slope of the peak power constrained capacity corresponds to $1 - 2f_d$, see [67], this unveils that the peak power constrained upper bound on the achievable rate in (3.78) is loose for $\beta > 1$ and high SNR.

On the other hand, the slope of the upper bound for proper Gaussian inputs given in (3.84) is equal to the slope of the lower bound as calculated in (3.109) following from the boundedness of the difference $\Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})}$, see (3.106). For this case the difference between the upper bound $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{\text{PG}}$ and the lower bound $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\text{PG}}$ converges to a constant for high SNR, cf. (3.107). Thus, both bounds must have the same asymptotic high SNR slope and we conjecture that the achievable rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})\big|_{\text{PG}}$ is also characterized by the same asymptotic SNR slope.

It is interesting to note that the high SNR slope of the capacity is degraded by the term $2f_d$. Now recall the discussion on the limits of the discrete-time input-output relation in Section 2.3. There it has been shown that symbol rate sampling does not yield a signal representation with a sufficient statistic as the normalized received signal bandwidth is given by $1 + 2f_d$. The excess bandwidth leading to aliasing is given by $2f_d$, which exactly corresponds

to the degradation of the high SNR slope of the achievable rate/capacity. Up to now, we do not know, if there is an implicit relation between these observations.

3.5 Comparison to Asymptotes in [67]

In [67], Lapidoth gives bounds for the peak power constrained capacity of non-coherent Rayleigh fading channels. These bounds are mainly derived to evaluate the asymptotic high SNR behavior. He distinguishes between two cases, *nonregular* and *regular* fading introduced by Doob [24]. The case of *nonregular* fading is characterized by the property that the prediction error variance of a one-step channel predictor — having infinitely many observations in the past — asymptotically approaches zero, when the SNR approaches infinity. As we consider the case that the PSD of the channel fading process is bandlimited with $f_d < 0.5$, our scenario corresponds to the *nonregular* case in [67], which is also named pre-log case. Fig. 3.4 shows a comparison of the lower bound on the capacity in (3.93)/(3.94) and the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in (3.84)/(3.87) with the high SNR asymptotes for the capacity in the corresponding pre-log case given in [67]. In contrast to our upper bound which holds only for proper Gaussian inputs, [67] does not constrain the input distribution except of a peak power constraint.

The capacity bounds in [67] are given by [67, eq. (33) and (47)]

$$C \leq \log \log \tilde{\rho} - \gamma - 1 + \log \left(\frac{1}{\epsilon_{pred}^2(1/\tilde{\rho})} \right) + o(1) \tag{3.111}$$

$$C \geq \log \left(\frac{1}{\epsilon_{pred}^2(4/\tilde{\rho}) + \frac{8}{5\tilde{\rho}}} \right) - \gamma - \log \left(\frac{1}{1 - \epsilon_{pred}^2(4/\tilde{\rho})} \right) - \log \left(\frac{5e}{6} \right), \tag{3.112}$$

where $\gamma \approx 0.577$ is the Euler constant and $\tilde{\rho}$ is defined as

$$\tilde{\rho} = \frac{P_{\text{peak}} \sigma_h^2}{\sigma_n^2} \tag{3.113}$$

i.e., it is an alternative definition of an SNR based on the peak power P_{peak} instead of the average power σ_x^2 used for the definition of the average SNR ρ in (2.18). Furthermore, $o(1)$ depends on the SNR and converges to zero for $\tilde{\rho} \to \infty$, i.e., $f(n) \in o(g(n))$ if

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0. \tag{3.114}$$

In addition, the prediction error variance $\epsilon_{pred}^2(\delta^2)$ is given by

$$\epsilon_{pred}^2(\delta^2) = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_h(f)}{\sigma_h^2} + \delta^2\right) df\right) - \delta^2. \qquad (3.115)$$

Although for the bounds on the peak power constrained capacity in [67] not an explicit average power constraint has been used, but only a peak power constraint, by this peak power constraint implicitly also a constraint on the average power is given. This should be obvious, as for the maximum average power σ_x^2 the inequality $\sigma_x^2 \leq P_{\text{peak}}$ must hold. Furthermore, it has to be considered that in case of using a peak power constraint, it is in general not optimal to use the maximum average power σ_x^2, see Section 3.4.1.1. In case the maximum average power σ_x^2 is not used, i.e., $\mathrm{E}\left[|x_k|^2\right] < \sigma_x^2$, the *nominal* SNR ρ as defined in (2.18) is not the actual average SNR. However, in the case of i.i.d. zero-mean proper Gaussian input symbols, the achievable rate is maximized when using the maximum average power σ_x^2, i.e., in this case the nominal average SNR ρ is also the actual average SNR.

As the peak power constraint that has been used for the bounds on the peak power constrained capacity in [67], i.e., for (3.111) and (3.112), implicitly constrains the average power to σ_x^2, for the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols and the bounds on the peak power constrained capacity in [67], we choose $\tilde{\rho}$ in (3.111) and (3.112) to be equal to the nominal average SNR ρ used for the bounds on the achievable rate with i.i.d. Gaussian input symbols, i.e., set $\sigma_x^2 = P_{\text{peak}}$.

Fig. 3.4 shows that the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., the lower bound in (3.93)/(3.94) and our upper bound in (3.84)/(3.87), are in between the asymptotes for the upper bound and the lower bound on capacity given in [67]. However, the bounds in [67] consider a peak power constrained input distribution. Therefore, this comparison is not absolutely fair. In addition, and this is the main observation from this comparison, our bounds have the same slope in the high SNR regime as the high SNR asymptotes for the peak power constrained capacity in [67].

3.6 Comparison to Bounds in [105] and [107]

In [105] and [107] capacity bounds for a Rayleigh flat-fading channel for the case of a peak power constraint are given. Notice, as these bounds are bounds on capacity, only the average and the peak power constraint are imposed in the input signal. Additional constraints, like i.d. input symbols, have not been used.

First, we want to compare the lower bounds on the capacity given in [105, (34)/(29)] and in (3.96). Therefore, first consider that in [105] no assumption on i.d. input symbols has been made. When discarding this assumption, we
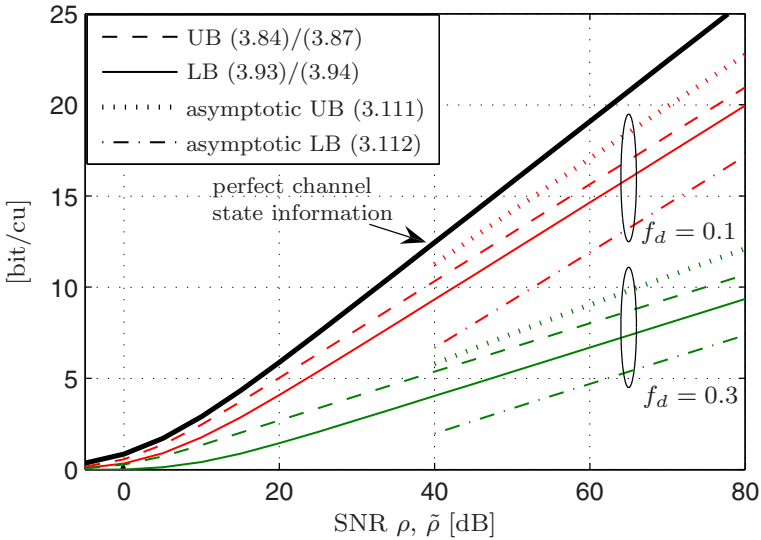
Fig. 3.4: Comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in (3.84)/(3.87) and (3.93)/(3.94) (SNR ρ) with asymptotic bounds on the peak power constrained capacity in (3.111) and (3.112) (SNR $\tilde{\rho}$), [67, eq. (33) and (47)] (The asymptotic upper bound (3.111) only holds for $\tilde{\rho} \to \infty$ as we neglect the term $o(1)$ in (3.111), which approaches zero for $\tilde{\rho} \to \infty$.); rectangular PSD of the channel fading process

can enhance the lower bound in (3.96), as the following argumentation will show. Obviously, it might be possible that we can get a larger lower bound on the capacity as the one in (3.96), if we use the channel only for a fraction of the time, and transmit nothing during the rest of the time, i.e., split the time into two blocks. In one of the blocks we use the channel, assuming i.d. input symbols, and in the other we transmit nothing. The ratio between the lengths of both blocks is held constant. Thus, in the limit, both blocks are still infinitely long. With this argumentation we deviate from the assumption of i.d. input symbols over the complete time horizon.

Using this time sharing argument, a lower bound on the peak power constrained capacity for input distributions with an average power σ_x^2 and a nominal peak-to-average power ratio β is consequently given by the following expression

$$\sup_{\mathcal{P}^{\mathrm{peak}}} \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) \geq \max_{\gamma \in [1,\beta]} \left\{ \frac{1}{\gamma} \mathcal{I}(y; x|h) \big|_{\mathrm{CM}, \gamma \sigma_x^2} - \frac{1}{\gamma} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left(\frac{\gamma \sigma_x^2}{\sigma_n^2} S_h(f) + 1 \right) df \right\}$$

$$(3.116)$$

where $\mathcal{P}^{\text{peak}}$ corresponds to $\mathcal{P}^{\text{peak}}_{\text{i.d.}}$ in (3.6) without the restriction to i.d. input symbols. The lower bound in [105, (34)/(29)] exactly corresponds to the lower bound in (3.116).[6]

In addition, the upper bound on the achievable rate with i.d. input symbols in (3.78) is a special case of the upper bound on the peak power constrained capacity given in [107, Proposition 2.2], which is on the one hand an upper bound on capacity as, except of the peak and average power constraints, no further assumptions on the input distributions have been made. On the other hand, the upper bound in [107, Proposition 2.2] holds for arbitrary PSDs of the channel fading process, while the derivation of the upper bound in (3.78) is based on the assumption of a rectangular PSD of the channel fading process. However, the approach of the derivation of the upper bound on the capacity given in [107] is completely different to our approach and is inherently based on the peak power constraint while we use this peak power constraint only in the last bounding step. Therefore, our lower bound on $h'(\mathbf{y}|\mathbf{x})$ in (3.60) also enables to give an upper bound on the achievable rate for non-peak power constrained input symbols.

As stated we made the assumption on identically distributed (i.d.) input symbols in the derivation of our upper bound. We do not know if this assumption poses a real restriction. Therefore, it would be necessary to know if the capacity achieving input distribution is characterized by identically distributed input symbols. But we have no answer to this question. However, as in case of a peak power constraint our upper bound on the achievable rate given in (3.78) corresponds to the upper bound on the peak power constrained capacity given in [107, Proposition 2.2], the restriction to identically distributed inputs seems not to be a severe restriction in the sense that it leads to an upper bound being lower than the capacity.

However, in [107] it is shown that i.i.d. inputs, i.e., with an additional constraint on independent input symbols, are not capacity-achieving in general. Based on the parameter

$$\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} |S_h(f)|^2 df \tag{3.117}$$

it has been shown in [107] that under the assumption of an absolutely summable autocorrelation function $r_h(l)$, see (2.4), in the asymptotic low SNR limit i.i.d. inputs are only capacity-achieving in the following two cases

- if $\lambda = \sigma_h^4$, corresponding to a memoryless channel,

---

[6] Note that it would also be possible to enhance the lower bound on the capacity for zero-mean proper Gaussian inputs in (3.93) based on the time sharing argument, i.e., by discarding the restriction to identically distributed input symbols. However, as for the derivation of the upper bound on the achievable rate in (3.84) we need the restriction to i.d. input symbols, such a lower bound without the assumption on i.d. input symbols would not match this upper bound. Therefore, we do not consider this further.

- or with a nominal peak-to-average power ratio of $\beta = 1$ and $\lambda \geq 2\sigma_h^4$, i.e., when the fading process is *nonephemeral*.

Notice that the proof in [107] explicitly is based on the asymptotic low SNR limit. On the other hand, for the high SNR case we have observed that i.i.d. zero-mean proper Gaussian inputs achieve the same asymptotic high SNR behavior in terms of the slope (pre-log) as the peak power constrained channel capacity, see Section 3.5.

3.7 Summary

In this chapter, we have derived bounds on the achievable rate with i.d. input symbols. The assumption on i.d. input symbols is required in the derivation of the upper bound on the achievable rate. We explicitly have given an upper bound for i.d. input symbols in combination with a peak power constraint, and on the other hand for i.i.d. zero-mean proper Gaussian input symbols.[7] Furthermore, we also give lower bounds on capacity, one for Gaussian inputs and one that is achievable with i.d. peak power constrained input symbols. The first one is already known from [21] and the second one is related to a lower bound on the peak power constrained capacity given in [105, (34)/(29)].

The main novelty of the present chapter lies in the new upper bounds on the achievable rate. These upper bounds are based on a new lower bound on the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ for the special case of a rectangular PSD of the channel fading process. This bound is not based on a peak power constraint, and therefore allows to give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in case of a rectangular PSD of the channel fading process. To the best of our knowledge, this is the only known upper bound on the achievable rate without a peak power constraint, which is tight in the sense that its slope (pre-log) corresponds to the slope of the lower bound on the capacity. In addition, based on the given lower bound on $h'(\mathbf{y}|\mathbf{x})$ we also have derived an upper bound on the achievable rate with i.d. peak power constrained input symbols, which is identified to be similar to an upper bound on capacity given in [107, Proposition 2.2]. However, for the derivation of our upper bound on the achievable rate we need the restriction to a rectangular PSD of the channel fading process whereas the upper bound on capacity given in [107, Proposition 2.2] holds for an arbitrary PSD of the channel fading process.

Furthermore, the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols with the asymptotic bounds on the peak power constrained capacity given in [67] shows the interesting

---

[7] In the context of proper Gaussian input symbols, we always use the term *i.i.d.* inputs, as i.i.d. zero-mean proper Gaussian input symbols are capacity achieving in the coherent scenario, and as our bounds also hold for i.i.d. input symbols.

fact that the achievable rate with i.i.d. zero-mean proper Gaussian inputs is characterized by the same asymptotic high SNR slope as the peak power constrained capacity. This shows that this kind of input distribution is not highly suboptimal with respect to its high SNR performance.

With the upper and lower bound on the achievable rate for i.i.d. zero-mean proper Gaussian inputs we have found a set of bounds, which is tight in the sense that their difference is bounded. We are able to bound this gap analytically by $\gamma(1 + 2f_d)$ [nat/cu] with the Euler constant $\gamma \approx 0.577$ and the maximum normalized Doppler frequency f_d. Thus, for the specific case of proper Gaussian inputs we give bounds, which are tight (in the sense given above) over the whole SNR range. In contrast, available bounds on capacity often focus only on a specific SNR range, e.g., [107] discusses the low SNR regime whereas [67] considers the high SNR regime.

The calculation of bounds on the achievable rate with non-peak power constrained input symbols has also the advantage that we can use a lower bound on the achievable rate which converges to the coherent capacity for asymptotically small channel dynamics. In contrast, lower bounds on the peak power constrained capacity will not show this behavior, as the coherent capacity is achieved by non-peak power constrained proper Gaussian input symbols.

Chapter 4
Bounds on the Achievable Rate of a Flat-Fading Channel Based on Prediction

In the previous chapter, we have derived bounds on the achievable rate of a discrete-time flat-fading channel and i.d. input symbols based on a purely mathematical derivation. In contrast to this, in the present chapter we derive upper bounds on the achievable rate with i.i.d. input symbols by expressing the mutual information rate based on the error variance of a one-step channel predictor. With respect to this, the effect of the channel dynamics, i.e., the PSD of the channel fading process, is captured by the one-step prediction error variance. This relation has already been shown in prior work, e.g., [67], [107], and [96]. Therefore, parts of the following calculations are closely related to derivations concerning the capacity of wideband OFDM systems in [96] as well as to [107]. While in [96] constant modulus input distributions have been considered, we extend this approach to the case of input distributions with varying magnitude. Therefore, we express the mutual information at an individual time instant based on the channel prediction error variance. The channel prediction error variance itself is a random quantity depending on the distribution of the past transmit symbols. To get to an upper bound on the achievable rate, we show that the calculation of the channel prediction error variance under the assumption that all past transmit symbols are constant modulus symbols, minimizes the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ over all i.i.d. input distributions fulfilling an average power constraint and, thus, can be used to upper-bound the achievable rate. As far as we know this argumentation is new. In contrast to the upper bound on the achievable rate given in the previous chapter, the upper bound given in the present chapter holds for channel fading processes with an arbitrary PSD with compact support. Furthermore, we try to give an enhanced upper bound on the channel output entropy rate $h'(\mathbf{y})$ based on channel power prediction. Unfortunately, we find no closed form solution for this upper bound. Instead, we will discuss the problems occurring in this bounding approach. Finally, we compare the upper bounds on the achievable rate based on the channel prediction error variance with the bounds given in Chapter 3. Parts of the present chapter have already been published in [27].

Concerning the assumptions on the input distribution in the present chapter we assume i.i.d. input symbols. In contrast, in the previous chapter we have only used the restriction to identically distributed input symbols, whereas independency of the input symbols has not been assumed. However, notice that the upper bounds on the achievable rate with i.d. input symbols given in Chapter 3 hold also for i.i.d. input symbols. Therefore, introducing the further restriction on independent input symbols in the present chapter, which is required for the following derivations, still allows the comparison of the bounds on the achievable rate that will be derived within the present chapter towards the bounds given in the previous chapter.

Furthermore, it is worth mentioning that the derivation of the bounds within the present chapter does not require the assumption on an absolutely summable autocorrelation function, see (2.4). I.e., the bounds that are derived in the present chapter hold for the larger class of square summable autocorrelation functions, in contrast to the upper bounds on the achievable rate derived in Chapter 3.

The assumption on i.i.d. input symbols requires the definition of the following two sets of input distributions: For the non-peak power constrained case

$$\mathcal{P}_{\text{i.i.d.}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, \quad p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i), \quad p(x_i) = p(x_j) \quad \forall i, j, \right.$$
$$\left. \mathrm{E}[|x_k|^2] \leq \sigma_x^2 \quad \forall k \right\} \tag{4.1}$$

and for the peak power constrained case

$$\mathcal{P}_{\text{i.i.d.}}^{\text{peak}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, \quad p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i), \quad p(x_i) = p(x_j) \quad \forall i, j, \right.$$
$$\left. \left\{ \mathrm{E}[|x_k|^2] \leq \sigma_x^2, \ p(x_k) = 0 \ \forall x_k : |x_k|^2 > P_{\text{peak}} \right\} \forall k \right\} \tag{4.2}$$

corresponding to the sets $\mathcal{P}_{\text{i.d.}}$ in (3.5) and $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$ in (3.6) with the additional restriction to independent input symbols.

As already described, the mutual information rate can be expressed by differential entropy rates as follows, cf. (3.10)

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}). \tag{4.3}$$

In the following, we give expressions or bounds for $h'(\mathbf{y})$ and $h'(\mathbf{y}|\mathbf{x})$ based on channel prediction.

4.1 Calculation of $h'(\mathbf{y}|\mathbf{x})$ based on Channel Prediction

First, we rewrite the entropy rate $h'(\mathbf{y}|\mathbf{x})$ based on the prediction error variance of the one-step predictor. Using the chain rule for differential entropy, we can express $h'(\mathbf{y}|\mathbf{x})$ as follows

$$
\begin{aligned}
h'(\mathbf{y}|\mathbf{x}) &= \lim_{N\to\infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x}) \\
&\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} h(y_k|\mathbf{x}, \mathbf{y}_1^{k-1}) \\
&\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} h(y_k|\mathbf{x}_1^k, \mathbf{y}_1^{k-1}) \\
&\overset{(c)}{=} \lim_{N\to\infty} h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})
\end{aligned}
\tag{4.4}
$$

where $\mathbf{y}_a^b$ indicates the subvector of $\mathbf{y}$ containing the elements from entry a to entry b. Equality (a) uses the chain rule for differential entropy, equality (b) uses the fact that y_k conditioned on $\mathbf{y}_1^{k-1}$ and $\mathbf{x}_1^k$ is independent of the symbols $\mathbf{x}_{k+1}^N$ due to the independency of the transmit symbols. Equality (c) follows from the ergodicity and stationarity of all processes, see also [19, Chapter 4.2].

As the following argumentation will show, the channel output y_N conditioned on $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$ is proper Gaussian and, thus, it is fully characterized by its mean and its variance. The mean can be given as

$$
\begin{aligned}
\mathrm{E}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] &= \mathrm{E}\left[x_N h_N + n_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\
&= x_N \mathrm{E}\left[h_N|\mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] \\
&= x_N \hat{h}_N
\end{aligned}
\tag{4.5}
$$

where $\hat{h}_N$ is the MMSE estimate of h_N based on the channel output observations at all previous time instances and the channel input symbols at these time instances. Thus, the estimate $\hat{h}_N$ is a deterministic function of the observations and the transmitted symbols at all time instances $1, \ldots, N-1$, i.e.,

$$
\hat{h}_N = \mathrm{E}\left[h_N|\mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right].
\tag{4.6}
$$

Based on $\hat{h}_N$ the channel output y_N can be written as

$$
y_N = x_N h_N + n_N = x_N \left(\hat{h}_N + e_N\right) + n_N
\tag{4.7}
$$

where e_N is the prediction error given by

$$
e_N = h_N - \hat{h}_N.
\tag{4.8}
$$

As both, the noise as well as the fading process, are jointly proper Gaussian, the MMSE estimate is equivalent to the linear minimum mean squared error (LMMSE). Thus, the estimate $\hat{h}_N$ is zero-mean proper Gaussian and due to the independency of the estimation error e_N and the estimate $\hat{h}_N$, it follows that the estimation error e_N is also zero-mean proper Gaussian.

As e_N is proper Gaussian, it can be easily seen by (4.7) that y_N conditioned on $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$ is also proper Gaussian. Thus, for the evaluation of $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$ we need to calculate the variance of the conditional channel output y_N which is given by

$$
\begin{aligned}
\operatorname{var}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] &= \mathrm{E}\left[\left|y_N - \mathrm{E}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right]\right|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\
&= \mathrm{E}\left[\left|y_N - x_N\hat{h}_N\right|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\
&= \mathrm{E}\left[\left|x_N(h_N - \hat{h}_N) + n_N\right|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\
&= |x_N|^2 \mathrm{E}\left[\left|h_N - \hat{h}_N\right|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] + \sigma_n^2 \\
&= |x_N|^2 \mathrm{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] + \sigma_n^2 \\
&= |x_N|^2 \sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1}) + \sigma_n^2
\end{aligned}
\tag{4.9}
$$

where

$$
\begin{aligned}
\sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1}) &= \mathrm{E}\left[\left|h_N - \hat{h}_N\right|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] \\
&= \mathrm{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] \\
&\stackrel{(a)}{=} \mathrm{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}\right]
\end{aligned}
\tag{4.10}
$$

is the prediction error variance of the MMSE estimator for $\hat{h}_N$. For (a) we have used the fact that the estimation error e_N is orthogonal and, thus, independent of the observations $\mathbf{y}_1^{N-1}$. Notice, the prediction error variance depends on the input symbols $\mathbf{x}_1^{N-1}$ which is indicated by writing $\sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1})$. As already stated, the estimate $\hat{h}_N$ is zero-mean proper Gaussian and its variance is given by

$$
\sigma_{\hat{h}}^2(\mathbf{x}_1^{N-1}) = \sigma_h^2 - \sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1}).
\tag{4.11}
$$

As y_N conditioned on $\mathbf{x}_1^N$ and $\mathbf{y}_1^{N-1}$ is proper Gaussian, using (4.9) and (4.10) its PDF is given by

$$p(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}) = \frac{\exp\left(-\frac{|y_N - \hat{h}_N x_N|^2}{|x_N|^2 \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) + \sigma_n^2}\right)}{\pi(|x_N|^2 \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) + \sigma_n^2)}$$

$$= p(y_N|\mathbf{x}_1^N, \hat{h}_N) \qquad (4.12)$$

i.e., the estimate $\hat{h}_N$ contains all information on h that is contained in $\mathbf{y}_1^{N-1}$ while knowing $\mathbf{x}_1^{N-1}$. With (4.12) we can use the following substitution

$$h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}) = h(y_N|x_N, \hat{h}_N, \mathbf{x}_1^{N-1}). \qquad (4.13)$$

Based on this MMSE channel predictor interpretation, we can rewrite the entropy $h(y_N|x_N, \hat{h}_N, \mathbf{x}_1^{N-1})$ as

$$h(y_N|x_N, \hat{h}_N, \mathbf{x}_1^{N-1}) = \mathrm{E}_{x_N}\left[\mathrm{E}_{\mathbf{x}_1^{N-1}}\left[\log\left(\pi e\left(\sigma_n^2 + \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})|x_N|^2\right)\right)\right]\right]. \qquad (4.14)$$

With (4.4), (4.13), and (4.14) we get for i.i.d. input symbols

$$h'(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(\pi e\left(\sigma_n^2 + \sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})|x_k|^2\right)\right)\right]\right] \qquad (4.15)$$

where $\sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})$ is the prediction error variance as given in (4.10) for an infinite number of channel observations in the past, i.e.,

$$\sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1}) = \lim_{N\to\infty}\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) = \lim_{N\to\infty}\mathrm{E}\left[|h_N - \hat{h}_N|^2\big|\mathbf{y}_1^{N-1}, \mathbf{x}_1^{N-1}\right] \qquad (4.16)$$

which is indicated by writing $\sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})$. Note that we have switched the notation and now predict at the time instant k instead of predicting at the time instant N with $N \to \infty$. This is possible, as the channel fading process is stationary, the input symbols are assumed to be i.i.d., and as we consider an infinitely long past. The channel prediction error variance $\sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})$ will be further discussed in Section 4.3.1.

4.2 Upper Bound on $h'(\mathbf{y})$

Besides the calculation of $h'(\mathbf{y}|\mathbf{x})$ based on the channel prediction error variance given in the previous section, we also want to discuss the channel output entropy rate $h'(\mathbf{y})$ based on a prediction approach. Pursuing this approach, we first show a simple alternative derivation of the upper bound on $h'(\mathbf{y})$ in (3.18), before we try to derive an enhanced upper bound on $h'(\mathbf{y})$ based on a prediction of the channel output power. However, this derivation does not

lead to a solution for an enhanced upper bound on $h'(\mathbf{y})$. Nevertheless, we show the approach and its inherent difficulties.

Based on the chain rule for mutual information, and using that the channel fading process is stationary and ergodic, and the assumption on i.i.d. input symbols, we get for the entropy rate of a stationary stochastic process [19, Chapter 4.2]

$$h'(\mathbf{y}) = \lim_{N \to \infty} h(y_N | \mathbf{y}_1^{N-1}). \tag{4.17}$$

4.2.1 Simple Upper Bound on $h'(\mathbf{y})$

A simple upper bound on $h'(\mathbf{y})$ is obviously given by the following derivation. As conditioning reduces entropy, we can upper-bound the RHS of (4.17) by

$$h(y_N | \mathbf{y}_1^{N-1}) \le h(y_N). \tag{4.18}$$

Using (4.18) and ergodicity and stationarity, we can upper-bound $h'(\mathbf{y})$ by

$$h'(\mathbf{y}) \le h(y_N) \overset{(a)}{\le} \log\left(\pi e\left(\alpha \sigma_x^2 \sigma_h^2 + \sigma_n^2\right)\right) = h_U'(\mathbf{y}) \tag{4.19}$$

where for (a) we used the fact that proper Gaussian distributions maximize entropy [82] and that the average transmit power is given by $\alpha \sigma_x^2$ with $\alpha \in [0,1]$ allowing for average transmit powers smaller than the maximum average transmit power σ_x^2, see Section 3.2.2. Obviously this upper bound is equal to the upper bound (3.18).

4.2.2 Ideas for an Enhanced Upper Bound on $h'(\mathbf{y})$

In this section, we give an approach for the derivation of an enhanced upper bound on the entropy rate $h'(\mathbf{y})$. In order to do so, we transform the entropy $h(y_N | \mathbf{y}_1^{N-1})$ in (4.17) as follows

$$h(y_N | \mathbf{y}_1^{N-1}) = h\left(x_N h_N + n_N | y_{N-1}, y_{N-2}, \ldots, y_1\right)$$
$$\overset{(a)}{=} h\left(x_N |h_N| + n_N | y_{N-1}, y_{N-2}, \ldots, y_1\right) \tag{4.20}$$

where (a) holds, as the phase of x_N is uniformly distributed and independent from symbol to symbol and, thus, the phase of y_N is independent from the phase of h_N. Hence, all information that is contained in $y_{N-1}, \ldots, y_1$ on y_N is about the magnitude $|h_N|$. The assumption on transmit symbols x_N with a uniformly distributed phase, i.e., the assumption that $p(x_N)$ is circularly

symmetric is not a severe restriction, as we are looking for an upper bound on the output entropy, and a uniform phase maximizes entropy.

Now, the idea is to find an upper bound on $h'(\mathbf{y})$ by constructing an estimator for $|h_N|$ based on $y_{N-1}, \ldots, y_1$ and evaluating its estimation error. As we aim to find an upper bound, it is not required that this estimator is optimal, e.g., in the sense that it corresponds to the minimum mean squared error estimate. Therefore, we use an estimator delivering an estimate $\widehat{|h_N|^2}$ of the power $|h_N|^2$.

The actual channel power $|h_N|^2$ can be expressed by

$$|h_N|^2 = \widehat{|h_N|^2} + \varepsilon_N \tag{4.21}$$

where the estimate

$$\widehat{|h_N|^2} = f(y_{N-1}, \ldots, y_1) \tag{4.22}$$

is a deterministic function on the prior channel outputs $y_{N-1}, \ldots, y_1$ and ε_N is the estimation error.

Introducing (4.21) in (4.20) yields

$$
\begin{aligned}
&h(x_N|h_N| + n_N|y_{N-1}, y_{N-2}, \ldots, y_1) \\
&= h\left(x_N\sqrt{\widehat{|h_N|^2} + \varepsilon_N} + n_N \,\middle|\, y_{N-1}, y_{N-2}, \ldots, y_1 \right) \\
&\leq h\left(x_N\sqrt{\widehat{|h_N|^2} + \varepsilon_N} + n_N \,\middle|\, \widehat{|h_N|^2} \right) \\
&= h\left(y_N \,\middle|\, \widehat{|h_N|^2} \right) \tag{4.23}
\end{aligned}
$$

where the last bound holds as $\widehat{|h_N|^2}$ cannot contain more information about $|h_N|$ than $y_{N-1}, \ldots, y_1$ due to the data processing inequality. In the next step, we upper-bound the RHS of (4.23) using the fact that proper Gaussian random variables are entropy maximizers [82]. Therefore, we use that y_N conditioned on $\widehat{|h_N|^2}$ is zero-mean, and its variance is given by

$$
\mathrm{E}\left[|y_N|^2 \,\middle|\, \widehat{|h_N|^2} \right] = \mathrm{E}\left[\left| x_N\sqrt{\widehat{|h_N|^2} + \varepsilon_N} + n_N \right|^2 \,\middle|\, \widehat{|h_N|^2} \right]
$$
$$
= \alpha\sigma_x^2\widehat{|h_N|^2} + \alpha\sigma_x^2\mathrm{E}[\varepsilon_N] + \sigma_n^2 \tag{4.24}
$$

with the average transmit power $\alpha\sigma_x^2$ and $\alpha \in [0, 1]$. Here we assume that the estimation error ε_N is independent of the observations $y_{N-1}, \ldots, y_1$ and thus also independent of the estimate $\widehat{|h_N|^2}$. With (4.24) the entropy rate $h'(\mathbf{y})$ is upper-bounded by

$$h'(\mathbf{y}) \le h'_{U_{\mathrm{enh}}}(\mathbf{y}) = \lim_{N \to \infty} \mathrm{E}_{\widehat{|h_N|^2}} \log\left(\pi e \left(\alpha\sigma_x^2 \widehat{|h_N|^2} + \alpha\sigma_x^2 \mathrm{E}[\varepsilon_N] + \sigma_n^2\right)\right).$$

$$(4.25)$$

In the following, we calculate the expectation of the prediction error $\mathrm{E}[\varepsilon_N]$ for the estimator defined subsequently. Our aim is to construct a channel predictor with the following two properties.

- The estimator is conditionally unbiased for asymptotically small channel dynamics. In addition, as a channel with asymptotic small channel dynamics can be estimated arbitrarily well, the following should hold

$$\lim_{f_d \to 0} \lim_{N \to \infty} \widehat{|h_N|^2} = |h_N|^2.$$

$$(4.26)$$

- Furthermore, in case of an uncorrelated channel, i.e., $r_h(l) = \sigma_h^2 \delta_l$, we aim to get $\mathrm{E}[\varepsilon_N]\big|_{f_d=0.5} = \sigma_h^2$ and, thus,

$$\mathrm{E}\left[\widehat{|h_N|^2}\right]\Big|_{f_d=0.5} = 0.$$

$$(4.27)$$

This behavior can be achieved by the following approach

$$\widehat{|h_N|^2} = \frac{1}{\alpha\sigma_x^2} \sum_{l=1}^{N-1} w_l \cdot \left(|y_l|^2 - \sigma_n^2\right)$$

$$(4.28)$$

where the filter coefficients

$$\mathbf{w} = [w_{N-1}, \ldots, w_1]^T$$

$$(4.29)$$

are calculated such that

$$\begin{aligned}
\widehat{|y_N|^2} &= \sum_{l=1}^{N-1} w_l \cdot \left(|y_l|^2 - \left(\sigma_n^2 + \alpha\sigma_x^2\sigma_h^2\right)\right) + \sigma_n^2 + \alpha\sigma_x^2\sigma_h^2 \\
&= \sum_{l=1}^{N-1} w_l \cdot \widetilde{|y_l|^2} + \sigma_n^2 + \alpha\sigma_x^2\sigma_h^2 \\
&= \widehat{\widetilde{|y_N|^2}} + \sigma_n^2 + \alpha\sigma_x^2\sigma_h^2
\end{aligned}$$

$$(4.30)$$

is the linear minimum mean squared error estimate (LMMSE) of $|y_N|^2$. Furthermore, notice that the mean of the observations $|y_k|^2$ is given by

$$\mathrm{E}\left[|y_k|^2\right] = \alpha\sigma_x^2\sigma_h^2 + \sigma_n^2$$

$$(4.31)$$

and we define $\widetilde{|y_k|^2}$ to be the zero-mean process corresponding to $|y_k|^2$, i.e.,

$$\widetilde{|y_k|^2} = |y_k|^2 - \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right). \tag{4.32}$$

Correspondingly, $\widehat{\widetilde{|y_N|^2}}$ is the LMMSE estimate of $\widetilde{|y_N|^2}$.

We now aim to calculate the mean of the estimation error ε_N, which is given by

$$\mathrm{E}\left[\varepsilon_N\right] = \mathrm{E}\left[|h_N|^2 - \widehat{|h_N|^2}\right]. \tag{4.33}$$

Using (4.28) and (4.30) the estimate $\widehat{|h_N|^2}$ of $|h_N|^2$ can be calculated based on an estimate $\widehat{|y_N|^2}$ of $|y_N|^2$ as follows

$$
\begin{aligned}
\widehat{|h_N|^2} &= \frac{1}{\alpha\sigma_x^2}\left(\widehat{|y_N|^2} - \sigma_n^2 - \alpha\sigma_x^2\sigma_h^2\left(1 - \sum_{l=1}^{N-1} w_l\right)\right)\\
&= \frac{1}{\alpha\sigma_x^2}\left(\widehat{\widetilde{|y_N|^2}} + (\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2) - \sigma_n^2 - \alpha\sigma_x^2\sigma_h^2\left(1 - \sum_{l=1}^{N-1} w_l\right)\right)\\
&= \frac{1}{\alpha\sigma_x^2}\left(\widehat{\widetilde{|y_N|^2}} + \alpha\sigma_x^2\sigma_h^2\sum_{l=1}^{N-1} w_l\right). \tag{4.34}
\end{aligned}
$$

Substituting (4.34) into (4.33) yields

$$
\begin{aligned}
\mathrm{E}\left[\varepsilon_N\right] &= \mathrm{E}\left[|h_N|^2 - \frac{1}{\alpha\sigma_x^2}\left(\widehat{\widetilde{|y_N|^2}} + \alpha\sigma_x^2\sigma_h^2\sum_{l=1}^{N-1} w_l\right)\right]\\
&= \sigma_h^2 - \frac{1}{\alpha\sigma_x^2}\mathrm{E}\left[\widehat{\widetilde{|y_N|^2}}\right] - \sigma_h^2\sum_{l=1}^{N-1} w_l\\
&\stackrel{(a)}{=} \sigma_h^2\left(1 - \sum_{l=1}^{N-1} w_l\right) \tag{4.35}
\end{aligned}
$$

where (a) results from the fact that $\widehat{\widetilde{|y_N|^2}}$ is zero-mean.

In Appendix A.4, we show that for an infinitely long past, i.e., $N \to \infty$, the mean of the error of the channel power prediction is given by

$$\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right] = \lim_{N\to\infty}\mathrm{E}\left[\varepsilon_N\right] = \sigma_h^2 H_{\mathrm{pred}}(0) \tag{4.36}$$

with

$$H_{\mathrm{pred}}(0) = \sqrt{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{S_{\widetilde{|y|^2}}(f)}{S_{\widetilde{|y|^2}}(0)}\right)df\right)} \tag{4.37}$$

where $S_{\widetilde{|y|^2}}(f)$ is the PSD of the process $\left\{\widetilde{|y|^2}\right\}$.

Furthermore, in Appendix A.4 closed form expressions for $\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right]$ are given for the special case that the channel fading process is characterized by a rectangular PSD, see (A.44) and (A.45).

Computation of the upper bound given in (4.25), requires to discuss also the distribution of the estimate $\widehat{|h_N|^2}$. And this poses the major unsolved problem in the derivation of an enhanced upper bound on $h'(\mathbf{y})$. We know that with (4.33) the power of $\widehat{|h_N|^2}$ is given by

$$\sigma_{\hat{h}}^2 = \mathrm{E}\left[\widehat{|h_N|^2}\right] = \sigma_h^2 - \mathrm{E}[\varepsilon_N]. \tag{4.38}$$

However, we are not able to give further statements on the PDF $p\left(\widehat{|h|^2}\right)$ and, thus, based on (4.25) we can only give the following non-closed form solution for the enhanced upper bound on $h'(\mathbf{y})$

$$h'(\mathbf{y}) \leq h'_{U_{\mathrm{enh}}}(\mathbf{y}) = \mathrm{E}_{\widehat{|h|^2}} \log\left(\pi e \left(\sigma_n^2 + \alpha\sigma_x^2 \mathrm{E}[\varepsilon_{\mathrm{pred}}] + \alpha\sigma_x^2 \widehat{|h|^2}\right)\right). \tag{4.39}$$

where $\mathrm{E}[\varepsilon_{\mathrm{pred}}]$ is given by (A.44), and (A.45).

Notice that further upper-bounding based on Jensen's inequality leads to the simple upper bound on $h'(\mathbf{y})$ given in (4.19).

However, it can easily be seen that for $f_d \to 0$ the upper bound $h'_{U_{\mathrm{enh}}}(\mathbf{y})$ becomes tight, as the mean of the estimation error $\mathrm{E}[\varepsilon_{\mathrm{pred}}]$ becomes zero and, thus, $h'_{U_{\mathrm{enh}}}(\mathbf{y})$ converges to $h'(\mathbf{y}|\mathbf{h})$. On the other hand, in case the channel changes arbitrary slowly, it can be estimated arbitrary well, hence $h'(\mathbf{y})$ approaches $h'(\mathbf{y}|\mathbf{h})$, showing that $h'_{U_{\mathrm{enh}}}(\mathbf{y})$ becomes tight for $f_d \to 0$. In contrast, this is not the case for $h'_U(\mathbf{y})$ given in (4.19).

4.3 Upper Bound on the Achievable Rate

Based on (4.3), (4.4), and (4.17) the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ can be expressed as

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty} \left\{h(y_N|\mathbf{y}_1^{N-1}) - h(y_N|\mathbf{y}_1^{N-1},\mathbf{x}_1^N)\right\}. \tag{4.40}$$

In the following, we will refer to the separation of mutual information given in (4.40) as channel *prediction separation*.

With (4.19) and (4.15), we can give the following upper bound on the achievable rate with i.i.d. input symbols

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \le \log\left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right) - \mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(\sigma_n^2 + \sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})|x_k|^2\right)\right]\right]$$

$$= \log\left(\alpha\rho + 1\right) - \mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(1 + \frac{\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})}{\sigma_n^2}|x_k|^2\right)\right]\right].$$

$$(4.41)$$

Obviously, the upper bound in (4.41) still depends on the channel prediction error variance $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ given in (4.16), which itself depends on the distribution of the input symbols in the past. Effectively $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ is itself a random quantity. For infinite transmission lengths, i.e., $N \to \infty$, its distribution is independent of the specific time instant k, as the channel fading process is stationary and as the transmit symbols are i.i.d..

In the following section we will upper-bound the RHS of (4.41) by maximizing the RHS of (4.41) over all i.i.d. input distributions of the past input symbols $\mathbf{x}_{-\infty}^{k-1}$ with an average power $\alpha\sigma_x^2$. Frankly speaking, we want to find the i.i.d. input distribution of the past transmit symbols, that yields a distribution of the channel prediction error variance $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ which maximizes the RHS of (4.41).

4.3.1 The Prediction Error Variance $\sigma_{e_{pred,\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$

The prediction error variance $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ in (4.16) depends on the distribution of the input symbols $\mathbf{x}_{-\infty}^{k-1}$. To construct an upper bound on the RHS of (4.41) we need to find a distribution of the transmit symbols in the past, i.e., $\mathbf{x}_{-\infty}^{k-1}$, which leads to a distribution of $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$, which maximizes the RHS of (4.41). Therefore, we have to express the channel prediction error variance $\sigma_{e_{\mathrm{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ as a function of the transmit symbols in the past, i.e., $\mathbf{x}_{-\infty}^{k-1}$. In a first step, we will give such an expression for the case of a finite past time horizon, i.e., for $\sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1})$ as given in (4.10). The channel prediction error variance in (4.10) can be expressed by

$$\sigma_{e_{\mathrm{pred}}}^2(\mathbf{x}_1^{N-1}) = \sigma_h^2 - \mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}^H \mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}^{-1} \mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} \qquad (4.42)$$

where $\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}$ is the correlation matrix of the observations $\mathbf{y}_1^{N-1}$ while the past transmit symbols $\mathbf{x}_1^{N-1}$ are known, i.e.,

$$\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}} = \mathrm{E}\left[\mathbf{y}_1^{N-1}(\mathbf{y}_1^{N-1})^H|\mathbf{x}_1^{N-1}\right]$$

$$= \mathrm{E}\left[\left(\mathbf{X}_{N-1}\mathbf{h}_1^{N-1} + \mathbf{n}_1^{N-1}\right)\left(\mathbf{X}_{N-1}\mathbf{h}_1^{N-1} + \mathbf{n}_1^{N-1}\right)^H|\mathbf{x}_1^{N-1}\right]$$

$$= \mathbf{X}_{N-1}\mathbf{R}_h\mathbf{X}_{N-1}^H + \sigma_n^2\mathbf{I}_{N-1} \qquad (4.43)$$

with $\mathbf{X}_{N-1}$ being a diagonal matrix containing the past transmit symbols such that

$$\mathbf{X}_{N-1} = \text{diag}\left(\mathbf{x}_1^{N-1}\right). \tag{4.44}$$

Furthermore, the vectors $\mathbf{h}_1^{N-1}$ and $\mathbf{n}_1^{N-1}$ contain the channel fading weights and the noise samples from time instant 1 to time instant $N-1$. In addition, $\mathbf{R}_h$ is the autocorrelation matrix of the channel fading process corresponding to (2.17), but of dimension $(N-1) \times (N-1)$, which is not explicitly stated.

The cross correlation vector $\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}$ between the fading weight h_N and the observation vector $\mathbf{y}_1^{N-1}$ while knowing the past transmit symbols $\mathbf{x}_1^{N-1}$ is given by

$$\begin{aligned}
\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} &= \mathrm{E}\left[\mathbf{y}_1^{N-1}h_N^*\big|\mathbf{x}_1^{N-1}\right] \\
&= \mathrm{E}\left[\left(\mathbf{X}_{N-1}\mathbf{h}_1^{N-1} + \mathbf{n}_1^{N-1}\right)h_N^*\big|\mathbf{x}_1^{N-1}\right] \\
&= \mathbf{X}_{N-1}\mathbf{r}_{h,\text{pred}} \tag{4.45}
\end{aligned}$$

with

$$\mathbf{r}_{h,\text{pred}} = \left[r_h(-(N-1))\ldots r_h(-1)\right]^T \tag{4.46}$$

where $r_h(l)$ is the autocorrelation function as defined in (2.2).

Substituting (4.43) and (4.45) into (4.42) yields

$$\begin{aligned}
\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) &= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\mathbf{X}_{N-1}^H\left(\mathbf{X}_{N-1}\mathbf{R}_h\mathbf{X}_{N-1}^H + \sigma_n^2\mathbf{I}_{N-1}\right)^{-1}\mathbf{X}_{N-1}\mathbf{r}_{h,\text{pred}} \\
&= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\left(\mathbf{X}_{N-1}^H\mathbf{X}_{N-1}\right)^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}} \\
&\stackrel{(a)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}} \tag{4.47}
\end{aligned}$$

where for (a) we have used the following substitution

$$\mathbf{Z} = \mathbf{X}_{N-1}^H\mathbf{X}_{N-1} \tag{4.48}$$

i.e., $\mathbf{Z}$ is a diagonal matrix containing the powers of the individual transmit symbols in the past from time instant 1 to $N-1$. For ease of notation we omit the index $N-1$.[1]

Remember that we want to derive an upper bound on the achievable rate with i.i.d. input symbols by maximizing the RHS of (4.41) over all i.i.d. distributions of the transmit symbols in the past with an average power $\alpha\sigma_x^2$.

---

[1] Note that the inverse of $\mathbf{Z}$ in (4.47) does not exist, if a diagonal element z_i of the matrix $\mathbf{Z}$ is zero, i.e., one transmit symbol has a zero power. However, as the representation of the prediction error variance in the first line of (4.47) shows, $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$ is continuous in $z_i = 0$ for all i and, thus, this does not lead to problems in the following derivation.

Obviously, the distribution of the phases of the past transmit symbols $\mathbf{x}_1^{N-1}$ has no influence on the channel prediction error variance $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$. Thus, it rests to evaluate, for which distribution of the power of the past transmit symbols the RHS of (4.41) is maximized. In the following, we will show that the RHS of (4.41) is maximized in case the past transmit symbols have a constant power $\alpha\sigma_x^2$. I.e., calculation of the prediction error variance under the assumption that the past transmit symbols are constant modulus symbols with transmit power $|x_k|^2 = \alpha\sigma_x^2$ maximizes the RHS of (4.41) over all i.i.d. input distributions for the given average power constraint of $\alpha\sigma_x^2$.

To prove this statement, we use the fact that the expression in the expectation operation at the RHS of (4.41) (but here for the case of a finite past time horizon), i.e.,

$$
\log\left(1 + \frac{\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})}{\sigma_n^2}|x_N|^2\right)
$$
$$
= \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right) \qquad (4.49)
$$

is convex with respect to each individual element of the diagonal of $\mathbf{Z}$, i.e., of $\mathbf{z} = \text{diag}(\mathbf{Z})$. The proof of the convexity of (4.49) is given in Appendix A.5. Based on this convexity, we can apply Jensen's inequality yielding

$$
\mathrm{E}_{\mathbf{z}}\left[\log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\right]
$$
$$
\geq \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\left(\mathrm{E}_{\mathbf{z}}\left[\mathbf{Z}\right]\right)^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)
$$
$$
= \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h + \frac{\sigma_n^2}{\alpha\sigma_x^2}\mathbf{I}_{N-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)
$$
$$
= \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\sigma_{e_{\text{pred,CM}}}^2\right) \qquad (4.50)
$$

where $\sigma_{e_{\text{pred,CM}}}^2$ is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with power $\alpha\sigma_x^2$. Here the index CM denotes constant modulus.

As this lower-bounding of the LHS of (4.50) can be performed for an arbitrary N, i.e., for an arbitrary long past, we can also conclude that

$$
\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(1 + \frac{\sigma_{e_{\text{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})}{\sigma_n^2}|x_k|^2\right)\right] \geq \log\left(1 + \frac{\sigma_{e_{\text{pred,CM}},\infty}^2}{\sigma_n^2}|x_k|^2\right)
$$
$$(4.51)$$

where $\sigma_{e_{\text{pred,CM}},\infty}^2$ is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with power $\alpha\sigma_x^2$ and an

infinitely long past observation horizon. In this case the prediction error variance is no longer a random quantity but is constant for all time instances k.

Thus, with (4.41) and (4.51) we can give the following upper bound on the achievable rate with i.i.d. transmit symbols

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \mathcal{I}'_U(\mathbf{y};\mathbf{x}) = \log\left(\alpha\rho + 1\right) - \mathrm{E}_{x_k}\left[\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}}{\sigma^2_n}|x_k|^2\right)\right].$$
(4.52)

Note that constant modulus input symbols are obviously in general not the capacity maximizing input distribution. However, we only use them to find a distribution of $\sigma^2_{e_{\mathrm{pred,\infty}}}(\mathbf{x}^{k-1}_{-\infty})$ that maximizes the RHS of (4.41).

In Appendix A.6 based on [129], we recall the known result that for constant modulus input symbols and an infinitely long observation interval the prediction error variance is given by, cf. [67]

$$\sigma^2_{e_{\mathrm{pred,CM,\infty}}} = \frac{\sigma^2_n}{\alpha\sigma^2_x}\left\{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(1 + \frac{\alpha\sigma^2_x}{\sigma^2_n}S_h(f)\right)df\right) - 1\right\}$$
(4.53)

For the special case of a rectangular PSD, (4.53) becomes

$$\sigma^2_{e_{\mathrm{pred,CM,\infty}}} = \frac{\sigma^2_n}{\alpha\sigma^2_x}\left\{\exp\left(2f_d\log\left(1 + \frac{\sigma^2_h\alpha\sigma^2_x}{2f_d\sigma^2_n}\right)\right) - 1\right\}.$$
(4.54)

Thus, with (4.52) and (4.53) we have found an upper bound on the achievable rate with i.i.d. input symbols.

4.3.2 Effect of Constraints on the Input Distribution

As in Chapter 3, we evaluate the upper bound given in (4.52) for different constraints on the input distribution. Following exactly the same argumentation as in Section 3.4.1.1 for the case of a peak power constraint, we get the following upper bound on the achievable rate with i.i.d. input symbols

$$\sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{i.i.d.}}} \mathcal{I}'_U(\mathbf{y};\mathbf{x})$$

$$= \sup_{\alpha\in[0,1]} \sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{i.i.d.}}|\alpha} \left\{\log\left(\alpha\rho + 1\right) - \mathrm{E}_{x_k}\left[\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}}{\sigma^2_n}|x_k|^2\right)\right]\right\}$$

$$= \sup_{\alpha\in[0,1]} \left\{\log\left(\alpha\rho + 1\right) - \frac{\alpha}{\beta}\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}}{\sigma^2_h}\rho\beta\right)\right\}$$
(4.55)

where it has to be considered that the prediction error variance $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}$ depends on α. Furthermore, the set $\mathcal{P}^{\mathrm{peak}}_{\mathrm{i.i.d.}}|\alpha$ corresponds to the set $\mathcal{P}^{\mathrm{peak}}_{\mathrm{i.i.d.}}$ with an average transmit power fixed to $\alpha\sigma^2_x$. Now, we would have to calculate the supremum of the RHS of (4.55) with respect to α which turns out to be difficult due to the dependency of $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}$ on α. However, $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}$ monotonically decreases with an increasing α. Furthermore, the RHS of (4.55) monotonically increases with a decreasing $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}$. Therefore, we can upper-bound the RHS of (4.55) by setting $\alpha = 1$ within $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}$ in (4.53), i.e., $\sigma^2_{e_{\mathrm{pred,CM,\infty}}}\big|_{\alpha=1}$, yielding

$$
\sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{i.i.d.}}} \mathcal{I}'_U(\mathbf{y};\mathbf{x}) \leq \sup_{\alpha\in[0,1]}\left\{\log(\alpha\rho+1) - \frac{\alpha}{\beta}\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}\big|_{\alpha=1}}{\sigma^2_h}\rho\beta\right)\right\}
$$

$$
= \log(\alpha_{\mathrm{opt}}\rho+1) - \frac{\alpha_{\mathrm{opt}}}{\beta}\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}\big|_{\alpha=1}}{\sigma^2_h}\rho\beta\right)
$$

$$
= \mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{pred},P_{\mathrm{peak}}} \tag{4.56}
$$

with

$$
\alpha_{\mathrm{opt}} = \min\left\{1, \left(\frac{1}{\beta}\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}\big|_{\alpha=1}}{\sigma^2_h}\rho\beta\right)\right)^{-1} - \frac{1}{\rho}\right\}. \tag{4.57}
$$

As the bound in (4.56) becomes loose for nominal peak-to-average power ratios $\beta \to \infty$, we also give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols which is given by

$$
\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{pred,PG}} = \log(\rho+1) - \int_{z=0}^{\infty}\log\left(1 + \frac{\sigma^2_{e_{\mathrm{pred,CM,\infty}}}\big|_{\alpha=1}}{\sigma^2_h}\rho z\right)e^{-z}dz \tag{4.58}
$$

where we set $\alpha = 1$, as in the non-peak power constrained case the upper bound is maximized for the maximum average transmit power σ^2_x.

As far as we know, this upper bound on the achievable rate is new. The innovation in the derivation of this bound lies in the fact that we separate the input symbols into the one at the time instant x_k and the previous input symbols contained in $\mathbf{x}^{k-1}_{-\infty}$. The latter ones are only relevant to calculate the prediction error variance. The prediction error variance itself is a random variable depending on the distribution of the past transmit symbols. To derive an upper bound on the achievable rate we have to find the i.i.d. input distribution of the transmit symbols with average power $\alpha\sigma^2_x$ which leads to a distribution of the prediction error variance which maximizes the upper bound on the achievable rate. We have shown that this maximum is given

for constant modulus input symbols, which led to a constant prediction error variance for all time instances.

As the assumption on constant modulus symbols is only used in the context of the prediction error variance, the upper bound on the achievable rate still holds for any i.i.d. input distribution with the given average power constraint. This allows us to evaluate this bound for the case of i.i.d. zero-mean proper Gaussian input symbols, leading to (4.58). To the best of our knowledge this argumentation is new.

4.4 Comparison to Bounds given in Chapter 3

In this section, we compare the bounds on the achievable rate with i.i.d. input symbols based on the channel prediction separation with the bounds on the achievable rate given in Chapter 3.

4.4.1 Numerical Evaluation

All following comparisons are for a rectangular PSD as the upper bounds on the achievable rate derived in Chapter 3 only hold for this type of PSD.

Fig. 4.1 shows the upper bound on the achievable rate with i.i.d. input symbols and a peak power constraint based on the channel prediction separation in comparison to the upper bound on the achievable rate given in Section 3.4. For both upper bounds we used in addition the combination with the coherent upper bound (3.87). For the case of a nominal peak-to-average power ratio $\beta = 1$, the upper bound given in Chapter 3 (3.78) and the prediction based upper bound (4.56) coincide. For $\beta = 2$, the prediction based upper bound is lower than the upper bound (3.78). Obviously, for the shown parameters the upper bound on the achievable rate based on prediction in (4.56) is tighter than the upper bound given in (3.78). At this point it is important to consider the following. The previous statement on tightness holds only when considering i.i.d. input symbols. As it has been stated in Section 3.6 the upper bound on the achievable rate in (3.78), which has been derived under the assumption of i.d. peak power constrained input symbols, is in terms of its expression equal to the upper bound on the peak power constrained capacity, i.e., without the restriction to i.d. input symbols, given in [107, Prop. 2.2]. On the other hand, the prediction based upper bound in (4.56) holds only for i.i.d. input symbols. The comparison of the prediction based upper bound on the achievable rate with i.i.d. input symbols to the lower bound on the peak power constrained capacity in [107, (32)], i.e., without any restriction to i.i.d. input symbols, which we recently have displayed in [27, Fig. 1], shows indeed that the prediction based upper bound on the
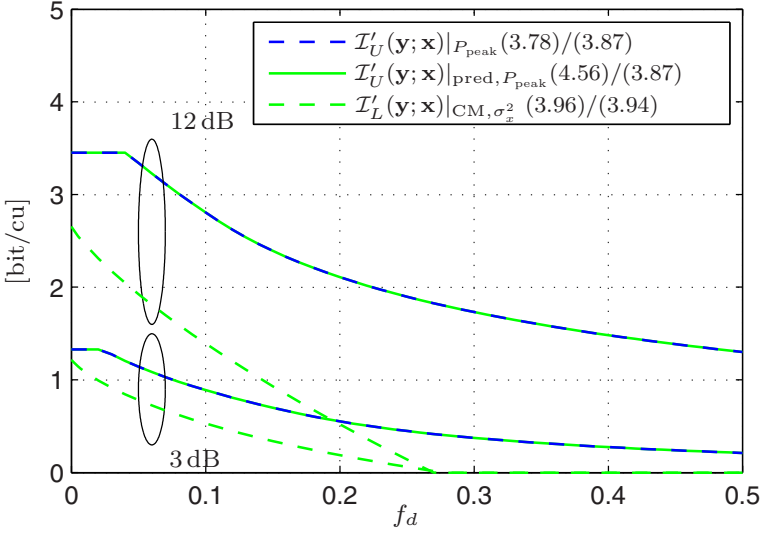
(a) $\beta = 1$
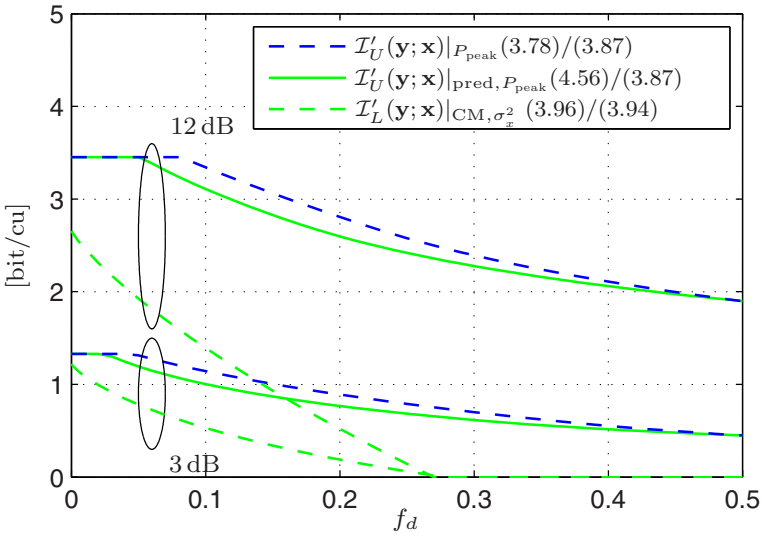
(b) $\beta = 2$

Fig. 4.1: **Comparison of the upper bound on the achievable rate with a peak power constraint** given in (4.56)/(3.87) based on channel prediction with the upper bound given in (3.78)/(3.87); in addition the lower bound on the peak power constrained capacity (3.96)/(3.94) is shown
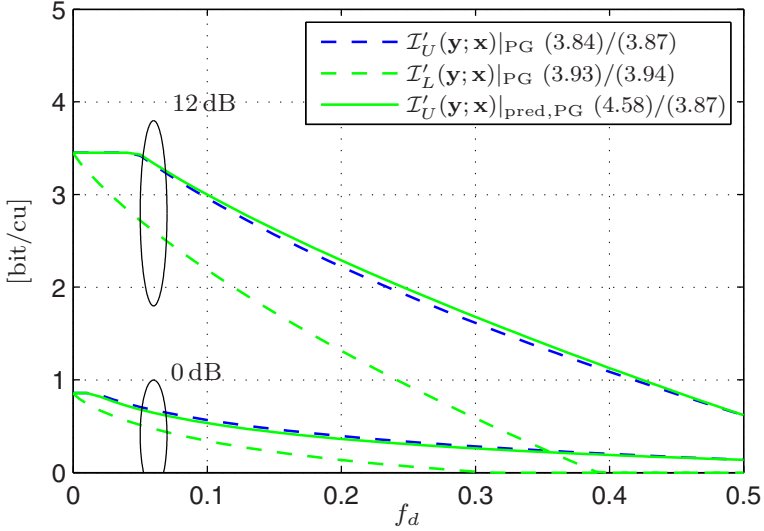
Fig. 4.2: **Comparison of the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols** based on the channel prediction separation (4.58)/(3.87) with the upper bound given in (3.84)/(3.87); in addition the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols (3.93)/(3.94) is shown

achievable rate with i.i.d. input symbols in (4.56) is not an upper bound on the peak power constrained capacity. This statement can be drawn, as the prediction based upper bound for i.i.d. input symbols in (4.56) is smaller than the lower bound on the peak power constrained capacity [107, (32)] for some choices of the system and channel parameters.

The prediction based upper bound (4.56) as well as (3.78) both become loose for $\beta > 1$ and high SNR or β very large, see Section 3.4.4 for (3.78). For comparison in Fig. 4.1 the lower bound on the peak power constrained capacity given in (3.96)/(3.94) is shown. However, this lower bound is in general not tight as it is based on a constant modulus input distribution.

Furthermore, Fig. 4.2 shows the prediction based upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (4.58) in comparison to the upper and lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs given in Section 3.4. As before, both upper bounds are shown in combination with the coherent upper bound (3.87).

Concerning the upper bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, a comparison of the prediction based upper bound (4.58) and the bound given in (3.84) shows, that it depends on the

channel parameters, which one is tighter. It can easily be shown that for $f_d \to 0$ and for $f_d = 0.5$ both bounds are equal. For other channel dynamics f_d it depends on the SNR ρ which bound is tighter. In Fig. 4.2 it can be observed that for the small SNR of 0 dB the prediction based upper bound (4.58)/(3.87) is smaller than (3.84)/(3.87) while for 12 dB the latter one is smaller and thus tighter. An analytical comparison turns out to be difficult as in both cases we use a very different way of lower bounding $h'(\mathbf{y}|\mathbf{x})$, which is further discussed in Section 4.4.2.

4.4.2 Relation of Bounds on $h'(\mathbf{y}|\mathbf{x})$

In this section, we discuss the relation between the bounds on the entropy rate $h'(\mathbf{y}|\mathbf{x})$ calculated in Section 3.3 and the lower bound on $h'(\mathbf{y}|\mathbf{x})$ based on the prediction error variance, i.e., (4.15) in combination with (4.53).

First, we consider the special case of constant modulus input symbols. Recall that for the special case of constant modulus input symbols the upper bound $h'_U(\mathbf{y}|\mathbf{x})$ in (3.29) corresponds to the actual value of $h'(\mathbf{y}|\mathbf{x})$, cf. (3.30). Thus, for constant modulus input symbols it holds that

$$h'(\mathbf{y}|\mathbf{x})\big|_{\mathrm{CM}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_h(f)\alpha\sigma_x^2}{\sigma_n^2} + 1\right) df + \log(\pi e \sigma_n^2). \tag{4.59}$$

On the other hand, the entropy rate $h'(\mathbf{y}|\mathbf{x})$ based on the channel prediction separation in (4.15) can only be given in closed form for constant modulus signaling, as otherwise no closed form expression for the prediction error variance $\sigma_{e_{\mathrm{pred}},\infty}^2(\mathbf{x}_{-\infty}^{k-1})$ is available. Assuming constant modulus signaling, $h'(\mathbf{y}|\mathbf{x})$ in (4.15) yields in combination with (4.53)

$$\begin{aligned}
h'(\mathbf{y}|\mathbf{x})\big|_{\mathrm{CM}} &= \lim_{N\to\infty} h(y_N | x_N, \hat{h}_N, \mathbf{x}_1^{N-1})\big|_{\mathrm{CM}} \\
&= \log\left(\pi e \left(\sigma_n^2 + \sigma_{e_{\mathrm{pred},\mathrm{CM}},\infty}^2 \alpha\sigma_x^2\right)\right) \\
&= \log\left(\pi e \left(\sigma_n^2 + \sigma_n^2 \left\{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(1 + \frac{\alpha\sigma_x^2}{\sigma_n^2} S_h(f)\right) df\right) - 1\right\}\right)\right) \\
&= \log\left(\pi e \sigma_n^2 \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(1 + \frac{\alpha\sigma_x^2}{\sigma_n^2} S_h(f)\right) df\right)\right) \\
&= \log\left(\pi e \sigma_n^2\right) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(1 + \frac{\alpha\sigma_x^2}{\sigma_n^2} S_h(f)\right) df \tag{4.60}
\end{aligned}$$

which exactly corresponds to (4.59).

This result is not surprising, as under the assumption of constant modulus inputs for both derivations no bounding has been used.

For the lower-bounding of $h'(\mathbf{y}|\mathbf{x})$ in case of a general input distribution, on the one hand in Section 3.3.2 inequality (3.58) has been used. On the other hand, for the bounds based on the channel prediction separation, we have shown that $h'(\mathbf{y}|\mathbf{x})$ is lower-bounded in case the prediction error variance $\sigma^2_{e_{\mathrm{pred},\infty}}(\mathbf{x}^{k-1}_{-\infty})$ is calculated based on constant modulus input symbols. An analytical comparison of these bounding techniques turns out to be difficult.

4.5 Summary

In the present chapter, we have derived an upper bound on the achievable rate with i.i.d. input symbols based on a prediction separation of mutual information. Based on this separation, the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ can be expressed by the one-step channel prediction error variance, which is a well known result, see e.g., [67], [105], and [107]. We show that for i.i.d. input symbols the prediction error variance $\sigma^2_{e_{\mathrm{pred},\infty}}(\mathbf{x}^{k-1}_{-\infty})$ calculated under the assumption of constant modulus symbols yields an upper bound on the achievable rate. As the constant modulus assumption is only used in the context of $\sigma^2_{e_{\mathrm{pred},\infty}}(\mathbf{x}^{k-1}_{-\infty})$, we can still give upper bounds on the achievable rate for general i.i.d. input symbol distributions, even for the case without a peak power constraint.

Furthermore, we have tried to give a new upper bound on the channel output entropy rate $h'(\mathbf{y})$ based on channel power prediction. Unfortunately this derivation does not lead to a closed form solution.

Comparing the upper bounds on the achievable rate with i.i.d. input symbols based on channel prediction calculated within the present chapter with the ones based on a pure mathematical derivation given in Chapter 3, it depends on the channel parameters which one is tighter. A further difference of the bounds lies in the fact that the ones given in Chapter 3 are based on a pure mathematical derivation, while the bounds in the present chapter give the link to a physical interpretation. Furthermore, for the derivation of the bounds in Chapter 3, the assumption of an absolutely summable autocorrelation function has been required, while this restriction is not necessary for the derivations in the present chapter. Finally, the upper bound given in the present chapter holds for channel fading processes with an arbitrary PSD of compact support, while the one given in Chapter 3 holds only for rectangular PSDs.

Chapter 5
Pilot Based Synchronized Detection

In this chapter, we compare the achievable rate with i.i.d. zero-mean proper Gaussian inputs as it has been considered in the two previous chapters to the achievable rate while using synchronized detection in combination with a solely pilot based channel estimation, as it is used in many typical receivers. In the context of synchronized detection, the receiver tasks are separated into two units. In the so-called inner receiver the unknown channel fading weights are estimated and then used for coherent detection/decoding[1] in the outer receiver.

In the first part of the present chapter, we shortly introduce the principle of synchronized detection. Afterwards, we recall existing bounds on the achievable rate while using synchronized detection in combination with a solely pilot based channel estimation given in [6] and [4]. In this context we also discuss, why a receiver using synchronized detection with a solely pilot based channel estimation cannot exploit the complete mutual information between transmitter and receiver. Finally, we compare these bounds to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in Chapter 3 to judge on the performance of pilot based synchronized detection with respect to the achievable rate.

5.1 Synchronized Detection

The task of the receiver is to decide on the most likely transmitted sequence $\mathbf{x}$ based on the observation of the channel output $\mathbf{y}$.[2] Relying on the discrete-time input-output relation given in (2.13), we can express this task in a mathematical way as

---

[1] We use the terms *detection* and *decoding* interchangeable.

[2] Notice that this corresponds to a receiver which aims to minimize the frame error rate.

$$\hat{\mathbf{x}} = \arg\max_{\mathcal{P}_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x}) \tag{5.1}$$

where $\mathcal{P}_{\mathbf{x}}$ is the set of all possible transmit sequences that can be generated by the transmitter. This set depends on the channel encoder and the modulation and is defined as the codebook. Obviously, (5.1) corresponds to *maximum likelihood* (ML) sequence detection. We assume here that all transmit sequences $\mathbf{x}$ in the set $\mathcal{P}_{\mathbf{x}}$ are generated by the transmitter with equal probability.

In case the probability of occurrence of the different sequences $\mathbf{x}$ is not equal, and the receiver has knowledge on $p(\mathbf{x})$, the receiver can make use of this knowledge by applying MAP detection instead, i.e.,

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\max_{\mathcal{P}_{\mathbf{x}}} p(\mathbf{x}|\mathbf{y}) \\ &= \arg\max_{\mathcal{P}_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \end{aligned} \tag{5.2}$$

However, we assume the transmitted sequences $\mathbf{x}$ to be equally likely and, thus, the MAP receiver in (5.2) becomes equivalent to the ML receiver in (5.1).

For zero-mean Gaussian fading channels, as introduced in Chapter 2, the conditional PDF $p(\mathbf{y}|\mathbf{x})$ is given by

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^N \det\left(\mathbf{R}_{y|x}\right)} \exp\left(-\mathbf{y}^H \mathbf{R}_{y|x}^{-1} \mathbf{y}\right) \tag{5.3}$$

with, cf. (3.20)

$$\mathbf{R}_{y|x} = \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N. \tag{5.4}$$

The evaluation of (5.1) therefore incorporates the calculation of a quadratic form and is prohibitively complex.

Based on a high SNR approximation, the solution of the optimization problem in (5.1) is approximatively equivalent to the solution of the following optimization problem [79, Section 12.2]

$$\begin{aligned}\left\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\right\} &= \arg\max_{\mathcal{P}_{\mathbf{x}},\mathbf{h}} p\left(\mathbf{h}|\mathbf{y}, \mathbf{x}\right) p(\mathbf{y}|\mathbf{x}) \\ &= \arg\max_{\mathcal{P}_{\mathbf{x}},\mathbf{h}} p\left(\mathbf{y}|\mathbf{x}, \mathbf{h}\right) p\left(\mathbf{h}\right) \end{aligned} \tag{5.5}$$

which is known as joint ML detection with MAP parameter estimation. For the special case of a jointly Gaussian problem as the one we discuss here, i.e., a proper Gaussian fading process and proper Gaussian noise, and constant modulus signaling the solution to the optimization problem in (5.5) is

exactly equivalent to the solution of the optimization problem in (5.1), see Appendix A.7 for a proof[3].

The solution space for $\hat{\mathbf{h}}$ is continuous whereas the solution space of $\hat{\mathbf{x}}$ is discrete. Therefore, the natural approach to the solution of (5.5) is to first maximize $p(\mathbf{h}|\mathbf{y}, \mathbf{x})$ with respect to $\mathbf{h}$ by MAP estimation for each possible sequence $\mathbf{x}$ and then choosing the sequence $\mathbf{x}$ with the largest likelihood, i.e., [79]

$$\hat{\mathbf{h}}(\mathbf{x}) = \arg \max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}, \mathbf{x}) \qquad (5.6)$$

$$\hat{\mathbf{x}} = \arg \max_{\mathcal{P}_{\mathbf{x}}} p\left(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})\right) p(\mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})). \qquad (5.7)$$

The notation $\hat{\mathbf{h}}(\mathbf{x})$ denotes that the estimate depends on the channel input symbols $\mathbf{x}$. Furthermore, writing $p(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}(\mathbf{x}))$ denotes that the channel estimate $\hat{\mathbf{h}}(\mathbf{x})$ is used as it would be the true value, see also (5.18).

Obviously, the first step corresponds to channel estimation, and the second step to detection/decoding based on the channel estimate $\hat{\mathbf{h}}$. For this approach the channel has to be estimated conditioned on all possible input sequences $\mathbf{x}$ which is prohibitively complex. Therefore, in conventional synchronized detection based receivers an approximative solution of (5.6) and (5.7) is achieved by performing the first step, i.e., the channel estimation based on pilot symbols, which are introduced into the transmit symbol sequence and are known to the receiver. Based on these channel estimates, the receiver can perform coherent detection/decoding of the data sequence. Fig. 5.1 shows the resulting structure.

5.1.1 Channel Estimation

For channel estimation, pilot symbols are multiplexed into the transmit symbol sequence. Based on the noisy channel observations at the pilot time instances, the channel fading process can be estimated by interpolation. For the Rayleigh fading model introduced in Chapter 2 the channel interpolation problem is jointly Gaussian. Hence, the MAP channel estimate in (5.6) corresponds to the MMSE estimate which is linear. Therefore, Wiener filtering delivers MMSE, and, thus, MAP channel estimates.

We assume that the pilot symbols are introduced periodically into the transmit symbol sequence with a pilot spacing L, i.e., each L-th symbol of the transmitted symbol sequence $\{x_k\}$ is a pilot symbol. To get a sufficient

---

[3] Note that in [79, Section 12.2.6 A] it has been stated that the solution of the optimization problem in (5.5) is equal to the solution of (5.1) for jointly Gaussian problems. There no restriction to constant modulus input distributions has been made. However, this equivalence does not hold in general as the proof in Appendix A.7 shows.
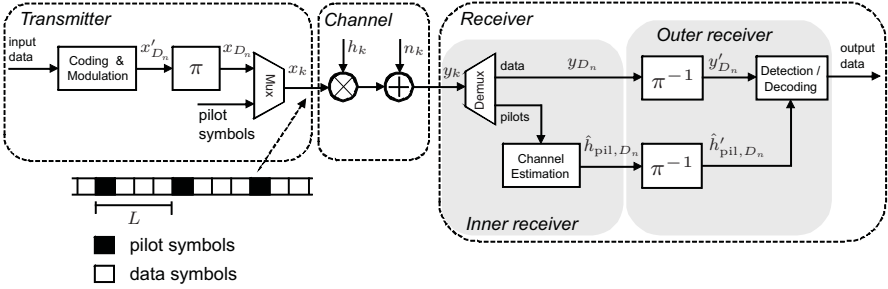
Fig. 5.1: **Block diagram of a synchronized detection receiver with a pilot based channel estimation**, flat-fading channel, simplified discrete-time baseband representation; $\pi/\,\pi^{-1}$ interleaving / deinterleaving; the index *pil* denotes that the channel estimate is solely based on pilot symbols and the index D_n indicates the n-th data symbol position

channel estimation quality, the pilot spacing L is chosen so that the channel fading process is sampled at least with Nyquist rate, i.e.,

$$L \leq \frac{1}{2f_d}. \tag{5.8}$$

The pilot symbols are constant modulus symbols, i.e., each having the transmit power σ_x^2. Note that thereby we deviate from the assumption that the channel input symbol stream consists of i.i.d. or i.d. symbols, as it has been used for the bounds on the achievable rate that have been calculated in the previous chapters.

Based on the pilot symbols, we estimate the channel fading process $\{h_k\}$, with $\{\hat{h}_{\mathrm{pil},k}\}$ being the channel estimation process. Here the index *pil* indicates that the channel estimation is solely based on pilot symbols. Based on the channel estimates $\hat{h}_{\mathrm{pil},k}$ the channel output observations y_k can be rewritten as

$$
\begin{aligned}
y_k &= h_k x_k + n_k \\
&= \left(\hat{h}_{\mathrm{pil},k} + e_{\mathrm{pil},k} \right) x_k + n_k \\
&= \hat{h}_{\mathrm{pil},k} x_k + e_{\mathrm{pil},k} x_k + n_k \\
&= \hat{h}_{\mathrm{pil},k} x_k + \tilde{n}_k
\end{aligned}
\tag{5.9}
$$

where $e_{\mathrm{pil},k} = h_k - \hat{h}_{\mathrm{pil},k}$ is the channel estimation error. The noise $\tilde{n}_k = e_{\mathrm{pil},k} x_k + n_k$ depends on the channel input x_k and is in general non-Gaussian. Based on the matrix-vector notation introduced in Section 2.2 the input-output relation in (5.9) can be expressed by

$$\mathbf{y} = \mathbf{X}\left(\hat{\mathbf{h}}_{\text{pil}} + \mathbf{e}_{\text{pil}}\right) + \mathbf{n} \tag{5.10}$$

where $\hat{\mathbf{h}}_{\text{pil}} = [\hat{h}_{\text{pil},1}, \ldots, \hat{h}_{\text{pil},N}]$ and $\mathbf{e}_{\text{pil}} = [e_{\text{pil},1}, \ldots, e_{\text{pil},N}]$.

Furthermore, we introduce the following subvectors $\mathbf{x}_P$ and $\mathbf{x}_D$, where the first one contains all pilot symbols contained in $\mathbf{x}$ and the latter one contains all data symbols of $\mathbf{x}$, i.e., the pilot symbol vector is given by

$$\mathbf{x}_P = \left[x_1, x_{L+1}, \ldots, x_{\lfloor \frac{N-1}{L} \rfloor L+1}\right] \tag{5.11}$$

where we assume without loss of generality that $\mathbf{x}$ starts with a pilot symbol. The observation vector at the pilot symbol time instances $\mathbf{y}_P$ and the channel weight vector at the pilot time instances $\mathbf{h}_P$ are defined accordingly.

As already stated the channel estimates $\hat{\mathbf{h}}_{\text{pil}}$ are given by LMMSE estimation, i.e.,

$$\hat{\mathbf{h}}_{\text{pil}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P\right]$$

$$= \mathbf{R}_{hh_P}\left(\sigma_x^2 \mathbf{R}_{h_P} + \sigma_n^2 \mathbf{I}_{\lfloor \frac{N}{L} \rfloor}\right)^{-1} \mathbf{X}_P^* \mathbf{y}_P \tag{5.12}$$

with

$$\mathbf{R}_{h_P} = \mathrm{E}\left[\mathbf{h}_P \mathbf{h}_P^H\right] \tag{5.13}$$

$$\mathbf{R}_{hh_P} = \mathrm{E}\left[\mathbf{h} \mathbf{h}_P^H\right] \tag{5.14}$$

and $\mathbf{X}_P = \text{diag}(\mathbf{x}_P)$. Notice, ignoring edge effects the LMMSE estimation in (5.12) can be realized by a set of L FIR filters.

Following similar ideas as in Section 4.1 it can be shown that the channel estimation error is zero-mean proper Gaussian, i.e., $e_{\text{pil},k} \sim \mathcal{CN}(0, \sigma_{e_{\text{pil}}}^2)$, see also [6]. Furthermore, in [6] it is shown that for an infinitely long observation interval the channel estimation error variance $\sigma_{e_{\text{pil}}}^2$ is independent of the time instant k due to the condition (5.8). However, it has to be kept in mind that the channel estimation error $e_{\text{pil},k}$ at different time instances is not independent, i.e., it is temporally correlated [90].

Given an infinitely long observation interval and a pilot spacing that fulfills the Nyquist condition (5.8), the channel estimation error variance $\sigma_{e_{\text{pil}}}^2$ can be calculated as

$$\sigma_{e_{\text{pil}}}^2 = \mathrm{E}\left[\left|h_k - \hat{h}_{\text{pil},k}\right|^2\right]$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{e_{\text{pil}}}(f) df$$

$$\stackrel{(a)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L} \frac{S_h(f)}{\sigma_h^2} + 1} df \tag{5.15}$$

where ρ specifies the SNR and is given in (2.18) and where $S_{e_{\mathrm{pil}}}(f)$ is the PSD of the channel estimation error process in case the channel estimation is solely based on pilot symbols. It is given in (A.144) in Appendix A.12 and has been used for (a). For the special case of a rectangular PSD $S_h(f)$ the estimation error variance is given by

$$\sigma^2_{e_{\mathrm{pil}}} = \frac{2f_d L}{2f_d L + \rho}\sigma^2_h. \tag{5.16}$$

In addition, the channel estimates are also zero-mean proper Gaussian, i.e., $\hat{h}_{\mathrm{pil},k} \sim \mathcal{CN}(0, \sigma^2_{\hat{h}_{\mathrm{pil}}})$ with

$$\sigma^2_{\hat{h}_{\mathrm{pil}}} = \sigma^2_h - \sigma^2_{e_{\mathrm{pil}}}. \tag{5.17}$$

5.1.2 Interleaving and Detection/Decoding

As shown in Fig. 5.1, the estimate $\hat{\mathbf{h}}_{\mathrm{pil},D}$, i.e., the elements of $\hat{\mathbf{h}}_{\mathrm{pil}}$ at the data symbol time instances are used for detection/decoding. Due to the temporal correlation of the fading process $\{h_k\}$, adjacent symbols are characterized by a similar channel attenuation. As typical channel encoders, e.g., convolutional codes, turbo codes, etc. are designed for white noise channels, i.e., are not able to correct burst errors very well, usually channel interleaving π and deinterleaving π^{-1} is used. Thus, the input to the detector/decoder is given by $\mathbf{y}'_D = \pi(\mathbf{y}_D)$. The deinterleaver breaks up the channel correlation. However, as the interleaving process corresponds to a permutation of the symbols, it can easily be seen that the interleaver only has the effect that the temporal correlation of the underlying channel fading process of adjacent symbols of the vector $\mathbf{y}'_D$, i.e., the input to the detector/decoder is broken up. This is important to achieve a sufficient coding gain. Concerning information theoretic interpretations, it should be kept in mind that the correlation of the elements of $\mathbf{y}'_D$ which have been adjacent before deinterleaving, i.e., in the vector $\mathbf{y}_D$, remains unchanged. Therefore, interleaving, which is a permutation, has no influence on the mutual information.

Now we discuss the detection/decoding, i.e., the step in (5.7). Without any approximations the metric corresponding to (5.7) is given by [79]

$$L_{\mathbf{x}} = \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}(\mathbf{x})\right)^H \mathbf{R}_n^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}(\mathbf{x})\right) + \left(\hat{\mathbf{h}}(\mathbf{x})\right)^H \mathbf{R}_h^{-1} \left(\hat{\mathbf{h}}(\mathbf{x})\right) \tag{5.18}$$

where $\mathbf{R}_n$ is the correlation matrix of the additive noise. As we assume white noise, it is diagonal with

$$\mathbf{R}_n = \sigma^2_n \mathbf{I}_N. \tag{5.19}$$

Due to the approximation to the solution of (5.6) and (5.7) by estimating the channel solely based on pilot symbols the channel estimate $\hat{\mathbf{h}}(\mathbf{x})$ is substituted by $\hat{\mathbf{h}}(\mathbf{x}_P) = \hat{\mathbf{h}}_{\mathrm{pil}}$, which is independent of the data symbols contained in $\mathbf{x}_D$. Furthermore, (5.7) translates into maximizing (5.18) over all possible data sequences $\mathbf{x}_D$. With this approximation and as the pilot sequence is deterministic, (5.6) and (5.7) become

$$\hat{\mathbf{h}}_{\mathrm{pil}} = \arg\max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P) \tag{5.20}$$

$$\hat{\mathbf{x}}_D = \arg\min_{\mathcal{P}_{\mathbf{x}}} \left\| \mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}} \right\|^2 \tag{5.21}$$

where we have used (5.19). The second term at the RHS of (5.18) can be neglected as it does not depend on $\mathbf{x}_D$.

Now we look at the problem from another perspective. We aim to perform detection while using the channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$ that is solely based on pilot symbols, see (5.20). We know that the channel estimation error, which is described by the vector $\mathbf{e}_{\mathrm{pil}}$, cf. (5.10), is zero-mean proper Gaussian with the following covariance matrix

$$\mathbf{R}_{e_{\mathrm{pil}}} = \mathrm{E}\left[\mathbf{e}_{\mathrm{pil}} \mathbf{e}_{\mathrm{pil}}^H \right]. \tag{5.22}$$

The diagonal elements of $\mathbf{R}_{e_{\mathrm{pil}}}$ are given by $\sigma_{e_{\mathrm{pil}}}^2$ in (5.15). As it already has been stated the channel estimation error process $\{e_{\mathrm{pil},k}\}$ is not white but temporally correlated, thus $\mathbf{R}_{e_{\mathrm{pil}}}$ is not diagonal. In this regard, see also the PSD corresponding to $\mathbf{R}_{e_{\mathrm{pil}}}$ which is given by $S_{e_{\mathrm{pil}}}(f)$ in (A.144) in Appendix A.12.

Performing ML sequence detection in this case would correspond to

$$\hat{\mathbf{x}}_D = \arg\max_{\mathcal{P}_{\mathbf{x}}} p\left(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}_{\mathrm{pil}} \right). \tag{5.23}$$

Obviously, $p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}_{\mathrm{pil}})$ is proper Gaussian and given by

$$p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}_{\mathrm{pil}}) = \frac{\exp\left\{ -\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}} \right)^H \left(\mathbf{X}\mathbf{R}_{e_{\mathrm{pil}}}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N \right)^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}} \right) \right\}}{\pi^N \det\left(\mathbf{X}\mathbf{R}_{e_{\mathrm{pil}}}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N \right)}. \tag{5.24}$$

Notice the difference between the first term of (5.18) and the metric corresponding to (5.24). The latter one accounts for the channel estimation error in the covariance matrix, i.e., $\mathbf{X}\mathbf{R}_{e_{\mathrm{pil}}}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N$, while the first one does not. Obviously, using a metric different to the one corresponding to (5.24), which considers the channel estimation error, leads to mismatch decoding.

However, it has been shown above that in the special case of constant modulus signaling the joint ML detection and MAP parameter estimation problem in (5.5) is no approximation and leads to the exact solution of the

genuine optimization problem in (5.1). Joint ML detection and MAP parameter estimation corresponds to evaluating (5.6) and (5.7), i.e., using the channel estimate as it would be the actual channel realization for detection. When studying (5.24) for constant modulus input symbols, this still leads to a different decoding metric than the one given in (5.21), i.e, when using the channel estimate as it would be the actual channel realization, as the matrix $\mathbf{R}_{e_{\mathrm{pil}}}$ is not diagonal, i.e., the channel estimation error is temporally correlated [90]. We assume that this difference arises from the fact that for the derivation of (5.21) we have assumed that the channel is estimated solely based on pilot symbols, which is an approximation to the genuine synchronized detection approach described by (5.6) and (5.7). Therefore, to get optimum ML decoding based on the pilot based channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$, we have to use a metric corresponding to (5.24).

As already stated at the beginning of this section, the decoder works with an interleaved observation sequence $\mathbf{y}'_D$. Using standard decoding techniques the temporal correlation of the estimation error cannot be exploited during decoding. Neglecting these temporal correlations we get the following approximation for (5.24)

$$p(\mathbf{y}|\mathbf{x},\hat{\mathbf{h}}_{\mathrm{pil}}) \approx \frac{\exp\left\{-\left(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}}\right)^H \left(\mathbf{X}\mathbf{X}^H\sigma_{e_{\mathrm{pil}}}^2 + \sigma_n^2\mathbf{I}_N\right)^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}}\right)\right\}}{\pi^N \prod_{k=1}^N \left(|x_k|^2\sigma_{e_{\mathrm{pil}}}^2 + \sigma_n^2\mathbf{I}_N\right)}.$$
(5.25)

The bounds on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation which we recall in the following section are based on this approximation.

Furthermore, and this is important to note, in the rest of this work, when using the term *synchronized detection*, we assume that the decoder uses a metric corresponding to (5.25). I.e., the channel estimation error variance is considered by the decoder, while the correlation of the channel estimation error process is not exploited. Obviously, with this usage of the term *synchronized detection* we deviate from its genuine definition, as due to considering the estimation error variance in the decoding metric, the channel estimate is not used as it would be the actual channel realization.

Definition 5.1. In this work, we use the term *synchronized detection* for receiver structures, where the channel is estimated and then used for subsequent decoding based on the metric given in (5.25). This means that the detector/decoder treats the estimation error process as additional additive *white* noise.

For completeness we additionally mention the following fact. Standard detectors/decoders are designed under the assumption that the noise variance is independent of the transmitted symbols. Therefore, a further approximation of the PDF in (5.25) is required. Approximation of the symbol powers

contained in $\mathbf{XX}^H$ by their expectation σ_x^2 leads to

$$p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}_{\text{pil}}) \approx \frac{1}{\pi^N \left(\sigma_x^2 \sigma_{e_{\text{pil}}}^2 + \sigma_n^2 \right)^N} \exp \left\{ -\frac{\left\| \mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\text{pil}} \right\|^2}{\sigma_x^2 \sigma_{e_{\text{pil}}}^2 + \sigma_n^2} \right\}. \tag{5.26}$$

Notice that for the special case of constant modulus input symbols the approximation step from (5.25) to (5.26) is not required as both are equal due to the fact that $|x_k|^2 = \sigma_x^2$. Based on (5.26) we get the following decoding metric

$$\hat{\mathbf{x}}_D = \arg\min_{\mathcal{P}_\mathbf{x}} \left\| \mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\text{pil}} \right\|^2 \tag{5.27}$$

which corresponds to (5.21). Obviously, when discarding the information contained in the temporal correlation of the channel estimation error, for the special case of constant modulus symbols, we get to the same ML detection metric, as it has been given for the case the detector uses the channel estimate as it would be the actual channel realization.

As this metric corresponds to a Euclidean distance which can be separated as follows

$$\left\| \mathbf{y} - \mathbf{X}\hat{\mathbf{h}}_{\text{pil}} \right\|^2 = \sum_{n=1}^{N} \left| y_n - x_n \hat{h}_{\text{pil},n} \right|^2 \tag{5.28}$$

standard decoders based on a symbol-wise detection/demapping, i.e., the Viterbi decoder or the MAP decoder, with a complexity that grows linear with N can be used.

In summary, the following important facts should be kept in mind: The genuine definition of synchronized detection states that the channel estimate is used as it would be the actual realization of the channel fading process. Deviating from this definition, we use the term *synchronized detection*, assuming that decoding is based on a metric, taking into account the channel estimation error variance but not its temporal correlation, see Definition 5.1. This corresponds to a symbol-wise detection, to which we sometimes refer to as *coherent detection*. The last assumption arises from the fact that usually decoding is performed after deinterleaving and typical decoders are not able to exploit the temporal correlation of the channel estimation error process. This fact is also kept by the bounds on the achievable rate that we recall in the following section.

In addition, the specific influence of the channel estimation error depending on the symbol powers $|x_k|^2$ is typically not used in detection/decoding. However, the bounds on the achievable rate in the following section are not based on this restriction and we also do not include this into the definition of the term synchronized detection given before.

5.2 Achievable Rate

In this section, we recall results on the achievable rate based on synchronized detection with a pilot based channel estimation as described in the previous section. These bounds are given in [6] and [4]. Before we discuss these bounds, we study the relation between the achievable rate with synchronized detection in combination with a solely pilot based channel estimation and the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ considered in the previous chapters. Therefore, we use the vectors $\mathbf{x}_D$ and $\mathbf{x}_P$ containing all data symbols, respectively all pilot symbols of $\mathbf{x}$. Accordingly we define $\mathbf{y}_P$, $\mathbf{y}_D$, $\mathbf{h}_D$, and $\hat{\mathbf{h}}_{\text{pil},D}$.

In the following, we assume that the data symbols are i.i.d.. Furthermore, for the discussion on the achievable rate with solely pilot based synchronized detection we do not use any peak power constraint, so that it is in general optimal to use the maximum average transmit power, i.e., $\mathrm{E}\left[|x_k|^2\right] = \sigma_x^2$, cf. (4.1). Consequently, the parameter ρ defined in (2.18) corresponds to the actual mean SNR.

Based on these definitions and assumptions, we are able to rewrite the mutual information between the channel input and the channel output as follows:

$$
\begin{aligned}
\mathcal{I}(\mathbf{y}; \mathbf{x}) &= \mathcal{I}(\mathbf{y}_P, \mathbf{y}_D; \mathbf{x}_P, \mathbf{x}_D) \\
&\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_D | \mathbf{x}_P) + \mathcal{I}(\mathbf{y}_P, \mathbf{y}_D; \mathbf{x}_P) \\
&\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_D, \mathbf{x}_P) - \mathcal{I}(\mathbf{x}_D; \mathbf{x}_P) + \mathcal{I}(\mathbf{y}_P, \mathbf{y}_D; \mathbf{x}_P) \\
&\overset{(c)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_D, \mathbf{x}_P) \\
&\overset{(d)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_D | \mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{x}_P) \\
&\overset{(e)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \mathbf{x}_P, \mathbf{y}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P | \mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{x}_P) \\
&\overset{(f)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \mathbf{y}_P, \mathbf{x}_P) \\
&\overset{(g)}{=} \mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\text{pil},D}, \mathbf{x}_P\right) \\
&\overset{(h)}{=} \mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\text{pil},D}\right) \quad\quad\quad\quad\quad\quad (5.29)
\end{aligned}
$$

where (a), (b), (d), and (e) are based on the chain rule for mutual information. For (c) we have used the fact that the pilot symbols contained in $\mathbf{x}_P$ are non-random, and, thus, $\mathcal{I}(\mathbf{x}_D; \mathbf{x}_P) = 0$ and $\mathcal{I}(\mathbf{y}_P, \mathbf{y}_D; \mathbf{x}_P) = 0$. Equality (f) is based on the fact that $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_P | \mathbf{x}_P) = 0$ as $\mathbf{y}_P$ does not carry any information on the data symbols $\mathbf{x}_D$, and we use again that $\mathcal{I}(\mathbf{x}_D; \mathbf{x}_P) = 0$ due to the fact that the pilot symbols are non-random. Equality (g) holds due to the fact that the channel estimate $\hat{\mathbf{h}}_{\text{pil},D}$ is a deterministic function of $\mathbf{y}_P$ and $\mathbf{x}_P$, and $\hat{\mathbf{h}}_{\text{pil},D}$ contains all information about $\mathbf{h}_D$ given by $\mathbf{y}_P$ while $\mathbf{x}_P$ is known. In Section 7.2 we also give a formal derivation of this

equality. As the pilot symbols are not only known, but also are deterministic, the conditioning on $\mathbf{x}_P$ can be ignored in (h). Nevertheless, in general, $\mathbf{x}_P$ has to be considered for the calculation of the correlation matrix of the channel estimation error $\mathbf{R}_{e_{\mathrm{pil}}}$ in (5.22). But for the special case of constant modulus pilot symbols, as they are assumed here, $\mathbf{R}_{e_{\mathrm{pil}}}$ does not explicitly depend on $\mathbf{x}_P$.

The derivation steps shown in (5.29) are not the direct way to get from $\mathcal{I}(\mathbf{y}; \mathbf{x})$ to $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D})$. However, the shown approach gives some additional insights. On the one hand, the mutual information $\mathcal{I}(\mathbf{y}; \mathbf{x})$, i.e., the mutual information between all channel input symbols and all channel observations is equal to the mutual information $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_D, \mathbf{x}_P)$, showing that the pilot symbols can also be understood as additional information at the receiver side. On the other hand, (5.29) gives the important result that $\mathcal{I}(\mathbf{y}; \mathbf{x})$ can be expressed by $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D})$, i.e., the mutual information between all input data symbols $\mathbf{x}_D$ and all channel output observations at the data symbol time instances $\mathbf{y}_D$, based on the side information $\mathbf{x}_P, \mathbf{y}_P$, i.e., the information on the channel delivered by the pilot symbols. This information is expressed by the estimate $\hat{\mathbf{h}}_{\mathrm{pil},D}$.

Obviously, $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D})$ is related to the achievable rate while using synchronized detection. The relation between $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D})$ and the achievable rate with synchronized detection is discussed in the following. More specifically, within the present chapter we discuss the case of synchronized detection with a solely pilot based channel estimation. In contrast, in Chapter 6, we will discuss the case of iterative code-aided synchronized detection, where reliability information on the data symbols is iteratively fed back to the channel estimator to enhance channel estimation quality, see Section 6.1.

Notice that the use of pilot symbols which are periodically multiplexed into the transmit symbol stream is not optimal in the sense of maximizing the mutual information $\mathcal{I}(\mathbf{y}; \mathbf{x})$. In Chapter 10, for PSK signaling the optimal distribution over the set of possible input sequences has been studied, where optimality is stated with respect to maximizing the achievable rate constrained to a fixed set of input sequences. This examination shows that the set of input distributions maximizing the achievable rate includes the use of a subset of all transmit sequences, where at one specific symbol position a fixed symbol is used. This means that for this scenario, the use of one single pilot symbol and of i.i.d. transmit symbols at all other time instances is optimal in the sense of maximizing mutual information. In contrast to this, we use pilot symbols which are periodically inserted into the transmit symbol stream. The advantage of this approach is that it enables a separation of channel estimation and decoding. This allows a receiver implementation with low complexity.

As it has already been stated, following the synchronized detection approach the channel estimate is used in the decoder corresponding to Definition 5.1. The channel estimation error is treated as additional noise, cf. (5.9). Considering the discussion on interleaving in Section 5.1.2, the temporal cor-

relation of the estimation error is not exploited by typical channel decoders. This means that the estimation error of adjacent input symbols/bits to the channel decoder is assumed to be independent. Thus, the RHS of (5.29), i.e., the term $\mathcal{I}(\mathbf{y}_D; \mathbf{x}_D | \hat{\mathbf{h}}_{\mathrm{pil},D})$ is only an upper bound on the achievable rate while using synchronized detection with a solely pilot based channel estimation. Therefore, we introduce the quantity $\mathcal{I}(\mathbf{y}_D; \mathbf{x}_D | \check{\hat{\mathbf{h}}}_{\mathrm{pil},D})$ where we assume that $\check{\hat{\mathbf{h}}}_{\mathrm{pil},D}$ is an estimate of the channel fading $\mathbf{h}_D$ with the same statistical properties as $\hat{\mathbf{h}}_{\mathrm{pil},D}$ except that the estimation error is temporally uncorrelated.[4]

Thus, $\mathcal{I}(\mathbf{y}_D; \mathbf{x}_D | \check{\hat{\mathbf{h}}}_{\mathrm{pil},D})$ is the mutual information while using synchronized detection with a solely pilot based channel estimation and a decoder which does not exploit the temporal correlation of the channel estimation error. Obviously, the following relation must hold

$$\mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right) \leq \mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \hat{\mathbf{h}}_{\mathrm{pil},D} \right.\right). \tag{5.31}$$

Assuming that the pilot spacing L fulfills the Nyquist channel sampling condition (5.8) leads in combination with an infinite observation interval to the fact that the channel estimation error variance is independent of the symbol time instant. Using this, the fact that the channel fading process is ergodic, and the assumption that the estimation error is temporally uncorrelated, the rate corresponding to $\mathcal{I}(\mathbf{y}_D; \mathbf{x}_D | \check{\hat{\mathbf{h}}}_{\mathrm{pil},D})$ is given by

$$\mathcal{I}'\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right) = \lim_{N \to \infty} \frac{1}{N} \mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right)$$
$$= \frac{L-1}{L} \mathcal{I}\left(y_{D_k}; x_{D_k} \left| \hat{h}_{\mathrm{pil},D_k} \right.\right) \tag{5.32}$$

where the index D_k indicates an arbitrarily chosen data symbol. The prefactor $\frac{L-1}{L}$ accounts for the loss of time instances that are used for pilot symbols and cannot be used for data transmission.

As the pilot symbols are deterministic, it is possible to use the following substitution

$$\mathcal{I}'\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right) = \mathcal{I}'\left(\mathbf{y}; \mathbf{x} \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil}} \right.\right) \tag{5.33}$$

which we apply in the following to get a notation being consistent with the rest of this work.

---

[4] Notice that it is not sufficient to substitute $\mathbf{y}_D$, $\mathbf{x}_D$, and $\hat{\mathbf{h}}_{\mathrm{pil},D}$ by the corresponding deinterleaved quantities $\mathbf{y}'_D$, $\mathbf{x}'_D$, and $\hat{\mathbf{h}}'_{\mathrm{pil},D}$ as the permutation by π has no influence on the mutual information, i.e.,

$$\mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \hat{\mathbf{h}}_{\mathrm{pil},D} \right.\right) = \mathcal{I}\left(\mathbf{y}'_D; \mathbf{x}'_D \left| \hat{\mathbf{h}}'_{\mathrm{pil},D} \right.\right). \tag{5.30}$$

In [77] and [6], bounds for $\mathcal{I}'\left(\mathbf{y};\mathbf{x}\middle|\check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)$ have been derived. We briefly summarize the bounding approach. Obviously, for the derivation of a lower bound on $\mathcal{I}'\left(\mathbf{y}_D;\mathbf{x}_D\middle|\check{\hat{\mathbf{h}}}_{\mathrm{pil},D}\right)$ a lower bound on $\mathcal{I}\left(y_{D_k};x_{D_k}\middle|\hat{h}_{\mathrm{pil},D_k}\right)$ in (5.32) is required. Therefore, the following separation is used

$$\mathcal{I}\left(y_{D_k};x_{D_k}\middle|\hat{h}_{\mathrm{pil},D_k}\right) = h(x_{D_k}|\hat{h}_{\mathrm{pil},D_k}) - h(x_{D_k}|y_{D_k},\hat{h}_{\mathrm{pil},D_k})$$

$$\overset{(a)}{=} h(x_{D_k}) - h(x_{D_k}|y_{D_k},\hat{h}_{\mathrm{pil},D_k}). \qquad (5.34)$$

where (a) is based on the independency of x_{D_k} and $\hat{h}_{\mathrm{pil},D_k}$. In [77] it is shown that the second term on the RHS of (5.34) is upper-bounded by

$$h(x_{D_k}|y_{D_k},\hat{h}_{\mathrm{pil},D_k}) \leq \mathrm{E}_{\hat{h}_{\mathrm{pil},D_k}} \log\left(\pi e \frac{\sigma_x^4 \sigma_{e_{\mathrm{pil}}}^2 + \sigma_n^2 \sigma_x^2}{\sigma_x^2|\hat{h}_{\mathrm{pil},D_k}|^2 + \sigma_x^2\sigma_{e_{\mathrm{pil}}}^2 + \sigma_n^2}\right). \qquad (5.35)$$

Furthermore, for zero-mean proper Gaussian data symbols the first term on the RHS of (5.34) becomes

$$h(x_{D_k}) = \log\left(\pi e \sigma_x^2\right). \qquad (5.36)$$

Inserting (5.35) and (5.36) into (5.34) and (5.32) results in the following lower bound on the achievable rate with solely pilot based synchronized detection and i.i.d. zero-mean proper Gaussian data symbols

$$\mathcal{I}'\left(\mathbf{y};\mathbf{x}\middle|\check{\hat{\mathbf{h}}}_{\mathrm{pil},D}\right) \geq \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\middle|\check{\hat{\mathbf{h}}}_{\mathrm{pil},D}\right)$$

$$= \frac{L-1}{L} \mathrm{E}_{\hat{h}_{\mathrm{pil},k}} \log\left(1 + \frac{\sigma_x^2|\hat{h}_{\mathrm{pil},k}|^2}{\sigma_{e_{\mathrm{pil}}}^2 \sigma_x^2 + \sigma_n^2}\right) \qquad (5.37)$$

$$= \frac{L-1}{L} \int_{z=0}^{\infty} \log\left(1 + \frac{\sigma_x^2(\sigma_h^2 - \sigma_{e_{\mathrm{pil}}}^2)}{\sigma_{e_{\mathrm{pil}}}^2 \sigma_x^2 + \sigma_n^2}z\right) e^{-z}dz \qquad (5.38)$$

$$= \frac{L-1}{L} \int_{z=0}^{\infty} \log\left(1 + \rho\eta z\right) e^{-z}dz \qquad (5.39)$$

where

$$\eta = \frac{1 - \frac{\sigma_{e_{\mathrm{pil}}}^2}{\sigma_h^2}}{1 + \frac{\sigma_{e_{\mathrm{pil}}}^2 \sigma_x^2}{\sigma_n^2}} \qquad (5.40)$$

is the SNR degradation factor.

On the other hand, in [6] it is shown that $\mathcal{I}'\left(\mathbf{y};\mathbf{x}\middle|\check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)$ can be upper-bounded by

$$\mathcal{I}'\left(\mathbf{y};\mathbf{x}\left|\check{\hat{\mathbf{h}}}_{\text{pil}}\right.\right) \leq \mathcal{I}'_U\left(\mathbf{y};\mathbf{x}\left|\check{\hat{\mathbf{h}}}_{\text{pil}}\right.\right)$$

$$= \frac{L-1}{L}\mathrm{E}_{\hat{h}_{\text{pil},k}}\log\left(1+\frac{\sigma_x^2|\hat{h}_{\text{pil},k}|^2}{\sigma_{e_{\text{pil}}}^2\sigma_x^2+\sigma_n^2}\right)+\frac{L-1}{L}\mathrm{E}_{x_k}\log\left(\frac{\sigma_x^2\sigma_{e_{\text{pil}}}^2+\sigma_n^2}{|x_k|^2\sigma_{e_{\text{pil}}}^2+\sigma_n^2}\right)$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\left|\check{\hat{\mathbf{h}}}_{\text{pil}}\right.\right)+\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})}. \tag{5.41}$$

The derivation of this upper bound is in principle based on the same approach as we will use it below in Section 6.2.5 in another context. For i.i.d. zero-mean proper Gaussian data symbols (5.41) becomes

$$\mathcal{I}'_U\left(\mathbf{y};\mathbf{x}\left|\check{\hat{\mathbf{h}}}_{\text{pil}}\right.\right)\bigg|_{\text{PG}}$$

$$= \frac{L-1}{L}\int_{z=0}^{\infty}\log\left(1+\frac{\sigma_x^2(\sigma_h^2-\sigma_{e_{\text{pil}}}^2)}{\sigma_{e_{\text{pil}}}^2\sigma_x^2+\sigma_n^2}z\right)e^{-z}dz$$

$$+ \frac{L-1}{L}\left(\log\left(\sigma_x^2\sigma_{e_{\text{pil}}}^2+\sigma_n^2\right)-\int_{z=0}^{\infty}\log\left(\sigma_x^2\sigma_{e_{\text{pil}}}^2 z+\sigma_n^2\right)e^{-z}dz\right)$$

$$= \frac{L-1}{L}\int_{z=0}^{\infty}\log\left(1+\frac{\sigma_x^2(\sigma_h^2-\sigma_{e_{\text{pil}}}^2)}{\sigma_{e_{\text{pil}}}^2\sigma_x^2+\sigma_n^2}z\right)e^{-z}dz$$

$$+ \frac{L-1}{L}\left(\log\left(\rho\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}+1\right)-\int_{z=0}^{\infty}\log\left(\rho\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}z+1\right)e^{-z}dz\right)$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\left|\check{\hat{\mathbf{h}}}_{\text{pil}}\right.\right)+\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})}\bigg|_{\text{PG}}. \tag{5.42}$$

By using Jensen's inequality, it is obvious that the term $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})}$ is non-negative and indicates the tightness of the upper and the lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})$. Furthermore, for $\sigma_{e_{\text{pil}}}^2 \to 0$, i.e., perfect channel knowledge, $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})}$ becomes 0. Regarding the bounding of $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x}|\check{\hat{\mathbf{h}}}_{\text{pil}})}|_{\text{PG}}$ notice its similarity to $\Delta_{h'(\mathbf{y})}$ in Section 3.4.3.2.

Effectively, these bounds show that for synchronized detection with a solely pilot based channel estimation, the achievable rate is decreased with respect to the case of perfect channel knowledge by two factors. First, the SNR is decreased by a factor η in (5.40) which arises because of an increased noise variance due to the estimation error and the fact that the variance of the channel estimate $\sigma_{\hat{h}_{\text{pil}}}^2$ is smaller than σ_h^2. Secondly, the achievable rate is decreased compared to the case of perfect channel knowledge by the factor $(L-1)/L$, as each L-th symbol is used for a pilot symbol and cannot be used for data transmission.

5.2.1 Comparison to the Achievable Rate with i.i.d. Gaussian Inputs

In this section, we compare the upper and the lower bound on the achievable rate with pilot based synchronized detection in (5.39) and (5.42) to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols that have been derived in Chapter 3.
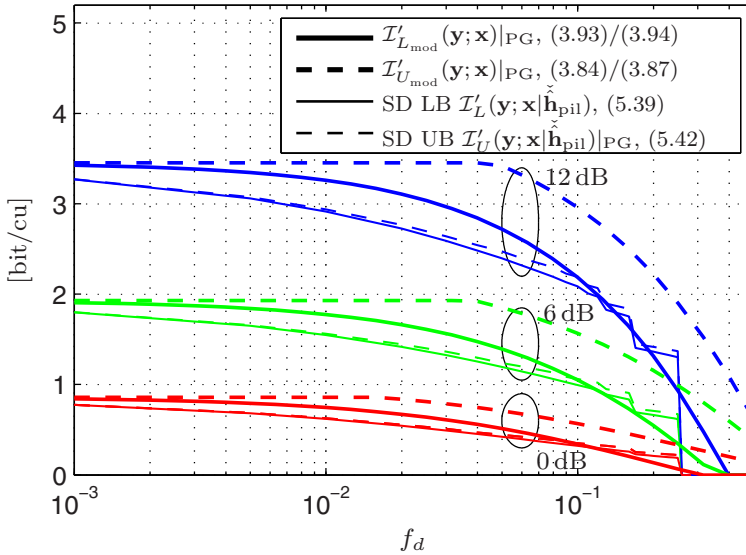
Fig. 5.2: Comparison of bounds on the achievable rate with pilot based synchronized detection (SD) to bounds on the achievable rate with i.i.d. input symbols; in both cases i.i.d. zero-mean proper Gaussian (data) symbols are assumed; rectangular PSD $S_h(f)$

Fig. 5.2 shows the bounds on the achievable rate using synchronized detection, (5.39) and (5.42), in comparison to the bounds on the achievable rate with i.i.d. input symbols, i.e., the lower bound (3.93)/(3.94) and the upper bound (3.84)/(3.87) for a rectangular PSD of the channel fading process. For the case of synchronized detection, we choose the pilot spacing L that maximizes (5.39). In both cases the data symbols are zero-mean proper Gaussian with an average power σ_x^2.

For $f_d \geq 0.25$, the achievable rate with synchronized detection is zero. This is due to the fact that in order to fulfill the Nyquist criterion, see (5.8), all symbols are pilot symbols. The unsteady behavior in the upper and lower bounds arise due to the fact that the pilot spacing can only be chosen as

an integer number. We can see that for low channel dynamics, i.e., small f_d, the achievable rate with synchronized detection is smaller than the lower bound on the achievable rate with i.i.d. input symbols, and, thus, lower than the channel capacity. This shows that with respect to the achievable rate, synchronized detection in combination with a solely pilot based channel estimation is suboptimal.

5.2.2 Optimized Pilot-to-Data Power Ratio

As discussed above, there exist two effects leading to a reduction of the achievable data rate while using synchronized detection in combination with a solely pilot based channel estimation compared to the case of perfect channel knowledge at the receiver side. On the one hand, this is the reduction in the number of data symbols depending on the pilot spacing L, which reduces the data rate by the factor $\frac{L-1}{L}$, and, on the other hand, the SNR loss due to the channel estimation error, which increases the data rate only logarithmically with increasing η. Therefore, from intuition, it is reasonable to make the pilot spacing as large as possible, while ensuring Nyquist sampling of the channel fading process by the pilot symbols, and to compensate the effect of the increased estimation error variance by optimizing the ratio between pilot and data power

$$\nu = \frac{\sigma_p^2}{\sigma_d^2} \tag{5.43}$$

where σ_p^2 is the power of the pilot symbols and σ_d^2 is the average power of the data symbols. This approach has already been evaluated in [4].

The maximum pilot spacing that can be chosen while fulfilling Nyquist sampling would be $L = \lfloor 1/(2f_d) \rfloor$. As here the pilot spacing is not a continuous function of f_d closed form evaluation becomes difficult. Therefore, in the following we use the approximation

$$L = \frac{1}{2f_d}. \tag{5.44}$$

For the important range of small f_d the difference in terms of the achievable rate is very small.

Writing both, the pilot power σ_p^2 and the average data power σ_d^2 as a function of ν and the overall average transmit power σ_x^2 yields

$$\sigma_p^2 = \frac{1}{2f_d + \frac{1}{\nu}(1 - 2f_d)}\sigma_x^2 \tag{5.45}$$

$$\sigma_d^2 = \frac{1}{1 - 2f_d + 2f_d\nu}\sigma_x^2. \tag{5.46}$$

The overall average transmit power σ_x^2 is held constant.

For a rectangular PSD $S_h(f)$ of the channel fading process and, thus, with $\sigma_{e_{\text{pil}}}^2$ in (5.16) (for the SNR of the pilot symbols) and with the pilot spacing in (5.44) the SNR degradation factor becomes

$$\eta = \frac{\rho\nu}{(\rho(1 + \nu) - 2f_d(1 - \nu) + 1)(2f_d(\nu - 1) + 1)}. \tag{5.47}$$

According to [4] the optimum choice of the power ratio is given by

$$\nu_{\text{opt}} = \sqrt{\frac{(1 - 2f_d)^2 + \rho(1 - 2f_d)}{2f_d(2f_d + \rho)}} \tag{5.48}$$

and due to Nyquist sampling of the channel fading process the pre-factor to the log becomes

$$\frac{L - 1}{L} = 1 - 2f_d. \tag{5.49}$$

Finally, the upper bound and lower bound on the achievable rate for synchronized detection and pilot power boosting are given by

$$\mathcal{I}'\left(\mathbf{y}; \mathbf{x} \middle| \check{\mathbf{h}}_{\text{pil}}\right) \geq \mathcal{I}_L'\left(\mathbf{y}; \mathbf{x} \middle| \check{\mathbf{h}}_{\text{pil}}\right)$$

$$= (1 - 2f_d) \int_{z=0}^{\infty} \log\left(1 + \eta\rho z\right) e^{-z} dz \tag{5.50}$$

$$\mathcal{I}'\left(\mathbf{y}; \mathbf{x} \middle| \check{\mathbf{h}}_{\text{pil}}\right) \leq \mathcal{I}_U'\left(\mathbf{y}; \mathbf{x} \middle| \check{\mathbf{h}}_{\text{pil}}\right)$$

$$= \mathcal{I}_L'\left(\mathbf{y}; \mathbf{x} \middle| \check{\mathbf{h}}_{\text{pil}}\right) + (1 - 2f_d)\mathrm{E}_{|x_D|^2} \log\left(\frac{\sigma_d^2\sigma_{e_{\text{pil}}}^2 + \sigma_n^2}{|x_D|^2\sigma_{e_{\text{pil}}}^2 z + \sigma_n^2}\right) \tag{5.51}$$

where x_D is an arbitrary data symbol with average power $\mathrm{E}\left[|x_D|^2\right] = \sigma_d^2$ and where, corresponding to (5.16), the estimation error variance is given by

$$\sigma_{e_{\text{pil}}}^2 = \frac{2f_d L}{2f_d L + \rho\frac{\sigma_p^2}{\sigma_x^2}}\sigma_h^2 = \frac{1}{1 + \rho\frac{\sigma_p^2}{\sigma_x^2}}\sigma_h^2 \tag{5.52}$$

as the SNR at the pilot symbol time instances is given by $\rho\frac{\sigma_p^2}{\sigma_x^2}$. For the special case of zero-mean proper Gaussian data symbols, (5.51) becomes

$$\mathcal{I}'_U\left(\mathbf{y};\mathbf{x}\left|\overset{\smash{\scriptscriptstyle\vee}}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right.\right)\Bigg|_{\mathrm{PG}} = \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\left|\overset{\smash{\scriptscriptstyle\vee}}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right.\right)$$

$$+ (1 - 2f_d)\int_{z=0}^{\infty}\log\left(\frac{\sigma_d^2\sigma_{e_{\mathrm{pil}}}^2 + \sigma_n^2}{\sigma_d^2\sigma_{e_{\mathrm{pil}}}^2 z + \sigma_n^2}\right)e^{-z}dz. \quad (5.53)$$
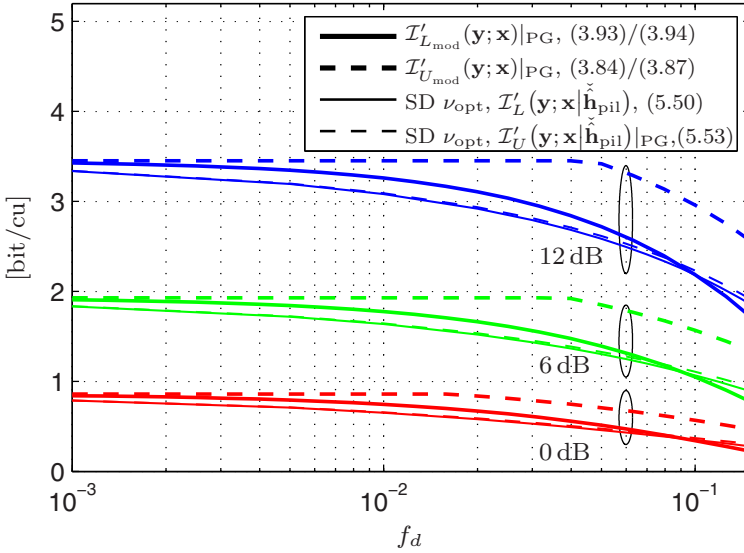
Fig. 5.3: Comparison of bounds on the achievable rate with pilot based synchronized detection with optimized pilot-to-data power ratio (SD ν_{opt}) to bounds on the achievable rate with i.i.d. input symbols; in both cases i.i.d. zero-mean proper Gaussian (data) symbols are assumed; approximation for the pilot spacing $L = 1/(2f_d)$; rectangular PSD $S_h(f)$

In Fig. 5.3 the achievable rate in case of using synchronized detection and pilot power-to-data power ratio optimization, see (5.50) and (5.53), is compared to the bounds on the achievable rate without pilot symbols, i.e., the lower bound (3.93)/(3.94) and the upper bound (3.84)/(3.87) for i.i.d. zero-mean proper Gaussian input symbols.

Obviously, the gap between the achievable rate with synchronized detection and pilot power optimization towards the lower bound on the achievable rate (3.93)/(3.94) is smaller compared to the case without pilot power optimization. However, for small f_d the achievable rate with synchronized detection and pilot power optimization is still lower than the lower bound (3.93)/(3.94), and thus lower as the channel capacity.

Nevertheless, the bounds using synchronized detection lie close to the lower bound on the achievable rate with an i.i.d. zero-mean proper Gaussian input distribution, indicating that the loss while restricting to synchronized detection in combination with a solely pilot based channel estimation is relatively small for low channel dynamics as they are typically observed in mobile channels.

5.3 Summary

In this chapter, we have recalled the principle of *synchronized detection* and have given a definition, how we use this term within the present work. This definition mainly states that synchronized detection means that the receiver estimates the channel allowing for coherent detection/decoding. The detector/decoder treats the channel estimation error as additive white noise, i.e., it discards the temporal correlation of the estimation error process. This is motivated by the fact that standard detectors/decoders do not exploit the information contained in the temporal correlation of the estimation error process.

In the second part of the present chapter, we have shown that the mutual information $\mathcal{I}(\mathbf{y};\mathbf{x})$ can be expressed in terms of the mutual information of the data symbols and the corresponding channel output observations conditioned on the information on the channel fading process delivered by the pilot symbols, which is expressed by $\hat{\mathbf{h}}_{\mathrm{pil},D}$, i.e., $\mathcal{I}\left(\mathbf{x}_D;\mathbf{y}_D\big|\hat{\mathbf{h}}_{\mathrm{pil},D}\right)$. We show that this is an upper bound on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation, which cannot be achieved as the detector/decoder does not exploit the temporal correlation of the channel estimation error.

Furthermore, we have recalled bounds on the achievable rate based on synchronized detection in combination with a solely pilot based channel estimation and compared them to the bounds on the achievable rate with i.i.d. input symbols, assuming zero-mean proper Gaussian data symbols in both cases. This comparison has shown that the achievable rate with pilot based synchronized detection stays below the achievable rate with i.i.d. input symbols for the very relevant case of small channel dynamics f_d. An exact evaluation of this gap is not possible as the bounds are not tight. However, for small channel dynamics this gap is not very large and becomes even smaller when using an optimized pilot-to-average data power ratio for the synchronized detection approach. For $f_d \to 0$ the bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation approach the coherent capacity.

Chapter 6
Iterative Code-Aided Synchronized Detection

In the previous chapter, we have derived bounds on the achievable rate using synchronized detection in combination with a solely pilot based channel estimation. In recent years receivers using iterative code-aided channel estimation got into the focus of research. The main idea behind this type of receivers is that—additionally to the pilot symbols which are used for an initial channel estimation and decoding—the channel estimation is enhanced by iteratively feeding back reliability information on the data symbols acquired by the channel decoder. Subsequently, this enhanced channel estimate is used in a further decoding step, permitting enhanced decoding results. We will refer to this type of receiver as *iterative code-aided synchronized detection* or *iterative synchronization and decoding* interchangeable. To evaluate the possible performance gain that can be achieved by receivers using iterative code-aided channel estimation and synchronized detection in comparison to receivers based on synchronized detection and a solely pilot based channel estimation, our aim is to study the achievable rate with such a type of receiver. For a specific type of such a receiver, which is a slight modification of the typically studied code-aided channel estimation based receiver, we give an upper bound on the achievable rate. This upper-bound is not a closed from expression. It explicitly depends on the channel interpolation error variance for an arbitrary time instant, and, thus, on all past and all future transmit symbols. However, for small channel dynamics, it is reasonable to approximate the channel interpolation error variance by the interpolation error variance calculated under the assumption that all past and future transmit symbols are constant modulus symbols. Based on this approximation we also numerically evaluate the upper bound.

As in case of the bounds on the achievable rate based on channel prediction given in Chapter 4, the derivations within the present chapter hold for square summable autocorrelation functions. I.e., the restriction to absolutely summable autocorrelation functions in (2.4) is not required.

Before starting with the derivation of this upper bound, we recall the principle of a receiver based on iterative code-aided channel estimation and

synchronized detection, i.e., iterative synchronization and decoding, and give an interpretation of the possible performance gain of this scheme in comparison to synchronized detection in combination with a solely pilot based channel estimation. Furthermore, we describe the modification of the typically studied iterative code-aided channel estimation based receiver, for which we afterwards derive an upper bound on the achievable rate.

6.1 Principle of Iterative Code-Aided Synchronized Detection

Here, we recall the principle of iterative code-aided synchronized detection. Fig. 6.1 shows the basic block diagram of a receiver following this principle. Before discussing the block-diagram in Fig. 6.1 we briefly recall the genuine
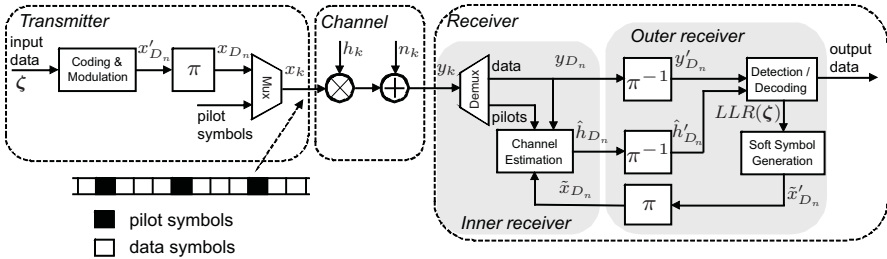
Fig. 6.1: **Block diagram of a receiver based on iterative code-aided synchronized detection**, SISO flat-fading channel; $\pi/\ \pi^{-1}$ interleaving / deinterleaving

principle of synchronized detection explained in Section 5.1 which can be formally expressed by the following set of equations, cf. (5.6) and (5.7)

$$\hat{\mathbf{h}}(\mathbf{x}) = \arg \max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}, \mathbf{x}) \tag{6.1}$$

$$\hat{\mathbf{x}} = \arg \max_{\mathcal{P}_{\mathbf{x}}} p\left(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})\right) p(\mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})). \tag{6.2}$$

Remember, (6.1) corresponds to the channel estimation task, whereas equation (6.2) expresses the detection/decoding task over the set of sequences contained in $\mathcal{P}_{\mathbf{x}}$ based on the assumption that the channel estimate $\hat{\mathbf{h}}(\mathbf{x})$ corresponds to the actual realization of the channel fading process $\mathbf{h}$. The notation $\hat{\mathbf{h}}(\mathbf{x})$ already shows that the estimate depends on the channel input sequence $\mathbf{x}$. In Chapter 5, we have restricted to the case that the channel estimate $\hat{\mathbf{h}}$ is solely based on deterministic pilot symbols, allowing for a sequential processing of channel estimation and detection/decoding. In the present sec-

tion, we consider the more general case, where also data symbols are used for channel estimation.

In typical code-aided synchronized detection based receivers—as they are recently studied in the literature, see, e.g., [125], [83], [41], [98], and [40]— both components, i.e., the channel estimator and the detection/decoding unit, are coupled iteratively, see Fig. 6.1. Typically, the first channel estimation is solely based on pilot symbols due to the lack of knowledge on the data symbols. Based on this channel estimate, the detector/decoder generates likelihood information on the transmitted sequence $\mathbf{x}$. Based on this likelihood information an enhanced channel estimate can be generated. This enhanced channel estimate itself is used in the subsequent detection/decoding step to get an enhanced estimate of $\mathbf{x}$, yielding updated reliability information on $\mathbf{x}$. This procedure is repeated several times.

In [98] it has been systematically shown that a corresponding receiver structure can be motivated by expressing the joint ML detection and MAP parameter estimation problem in (5.5) based on a set of fixed point equations. By iteratively solving the set of fixed point equations, the algorithm converges to one fixed point. In general there exists not necessarily only one fixed point, i.e., one solution to the set of fixed point equations, see [29] for a corresponding discussion on Turbo decoding. However, one solution of this set of fixed point equations corresponds to the solution of the genuine optimization problem in (5.5). This solution will be termed *correct fixed point* in the following.[1] The iterative solution of the set of fixed point equations motivates the iterative code-aided synchronized detection approach. In the following, we briefly recall the derivation of this approach, which is given in [98, Chapter 7]. The same receiver structure can also be derived based on the expectation maximization (EM) framework [20], see [40].

We start with the joint optimization problem in (5.5)

$$\left\{ \hat{\mathbf{x}}, \hat{\mathbf{h}} \right\} = \arg \max_{\mathcal{P}_{\mathbf{x}}, \mathbf{h}} p\left(\mathbf{y} | \mathbf{x}, \mathbf{h}\right) p\left(\mathbf{h}\right) \tag{6.3}$$

which is an high SNR approximation to the genuine ML receiver given in (5.1), and which can be shown to be optimal in case of a jointly Gaussian problem, i.e., jointly proper Gaussian fading and additive Gaussian noise, and constant modulus signaling, see the discussion in Section 5.1 and the proof in Appendix A.7. In a first step, in [98] the optimization problem in (6.3) is rewritten as the following optimization problem

$$\left\{ \hat{\boldsymbol{\lambda}}_I, \hat{\mathbf{h}} \right\} = \arg \max_{\boldsymbol{\lambda}_I, \mathbf{h}} p\left(\mathbf{y} | \boldsymbol{\lambda}_I, \mathbf{h}\right) p\left(\mathbf{h}\right) \tag{6.4}$$

where $\boldsymbol{\lambda}_I$ is a vector containing the reliability information of the information bits corresponding to the transmitted sequence $\mathbf{x}$. There exists a bijec-

---

[1] Note that the correct fixed point does not necessarily correspond to the correct code word.

tive mapping of $\boldsymbol{\lambda}_I$ on $\mathbf{x}$. This means that the discrete detection/decoding problem is transferred to a continuous parameter estimation problem. The optimization problem in (6.4) can be expressed by a set of fixed point equations [98]. The principle is briefly summarized in the following. This is mainly done for the channel estimator, which is moreover slightly rewritten to consider not only BPSK modulation but arbitrary signal constellations. For the detection/decoding, we will abstract from the presentation in [98] and do not consider the soft-information exchange within the detection/decoding unit, i.e., between the soft-demapper and the MAP-decoder. We only discuss the detection metric, which is sufficient in the context of the present work, such that detection/decoding is presented by a descriptor giving the reliability information delivered by the decoder.

The log-likelihood function corresponding to (6.4) is given by

$$L(\boldsymbol{\lambda}_I, \mathbf{h}) = \log\left(p(\mathbf{y}|\boldsymbol{\lambda}_I, \mathbf{h})\right) + \log\left(p(\mathbf{h})\right). \tag{6.5}$$

For a solution of the optimization problem in (6.4) the maximum of (6.5) has to be determined. Obviously, the following conditions are necessary for a local maximum of $L(\boldsymbol{\lambda}_I, \mathbf{h})$

$$\frac{\partial}{\partial \boldsymbol{\lambda}_I} L(\boldsymbol{\lambda}_I, \mathbf{h}) = \mathbf{0} \tag{6.6}$$

$$\frac{\partial}{\partial \mathbf{h}} L(\boldsymbol{\lambda}_I, \mathbf{h}) = \mathbf{0}. \tag{6.7}$$

This leads to the following expression for $\mathbf{h}$

$$\mathbf{h} = \left\{ \sum_{\mathbf{x}_i \in \mathcal{P}_x} p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}) \frac{1}{\sigma_n^2} \mathbf{X}_i \mathbf{X}_i^H + \mathbf{R}_h^{-1} \right\}^{-1} \sum_{\mathbf{x}_i \in \mathcal{P}_x} p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}) \frac{1}{\sigma_n^2} \mathbf{X}_i^* \mathbf{y} \tag{6.8}$$

where $p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h})$ gives the probability of the different sequences $\mathbf{x}_i$ based on the soft-information $\boldsymbol{\lambda}_I$ delivered by the decoder. Equation (6.8) can be identified as a channel estimator.

For the special case of constant modulus input symbols (6.8) simplifies to

$$\mathbf{h} = \left\{ \frac{\sigma_x^2}{\sigma_n^2} \mathbf{I}_N + \mathbf{R}_h^{-1} \right\}^{-1} \sum_{\mathbf{x}_i \in \mathcal{P}_x} p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}) \frac{1}{\sigma_n^2} \mathbf{X}_i^* \mathbf{y}$$

$$= \mathbf{R}_h \left\{ \mathbf{R}_h + \frac{\sigma_n^2}{\sigma_x^2} \mathbf{I}_N \right\}^{-1} \frac{1}{\sigma_x^2} \tilde{\mathbf{X}}^* \mathbf{y} \tag{6.9}$$

with $\tilde{\mathbf{X}}$ being a diagonal matrix containing the soft-symbols $\tilde{x}_k$, given by

$$\left[\tilde{\mathbf{X}}\right]_{kk} = \tilde{x}_k = \sum_{\mathbf{x}_i \in \mathcal{P}_x} p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}) x_{i,k}. \tag{6.10}$$

Obviously, for this special case of constant modulus input symbols the channel estimator has the structure of an LMMSE estimator. Thus, the channel estimates are calculated based on filtering of the soft-symbols.

The soft-information is contained in the PDFs $p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h})$, which will be delivered by the decoder. Thus, based on the current soft-information on the transmitted sequence $\mathbf{x}$ expressed by the soft-symbols $\tilde{x}_k$, we can get an enhanced estimate of the channel fading vector $\mathbf{h}$, which is enhanced with respect to the first purely pilot symbol based channel estimate.

Now, we turn our attention to the detector/decoder. The set of fixed point equations given in [98] consists, on the one hand, of the equation describing the channel estimator given in (6.8) and, on the other hand, of fixed point equations for all code bits describing detection/decoding. Due to the processing of soft-information detection/decoding corresponds to a soft-demapper followed by a MAP-decoder. To get the equations of the fixed point system corresponding to the detection/decoding, we would have to differentiate the log-likelihood function in (6.5) with respect to the soft-information $\boldsymbol{\lambda}_I$. As initially stated we prescind from this here and depict detection/decoding based on a description function, indicating the used detection metric, and delivering the reliability information on the transmitted sequence, i.e.,

$$\left\{ p\left(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}\right) \right\} = \mathrm{Dec}\left(p(\mathbf{y}|\mathbf{x}, \mathbf{h}), \forall \mathbf{x}_i \in \mathcal{P}_\mathbf{x} \right) \tag{6.11}$$

where the set $\mathcal{P}_\mathbf{x}$ contains all possible transmit sequences $\mathbf{x}_i$. Corresponding to (6.3), the PDF $p(\mathbf{y}|\mathbf{x}, \mathbf{h})$ is given by

$$p(\mathbf{y}|\mathbf{x}, \mathbf{h}) = \frac{1}{\pi^N \sigma_n^2} \exp\left(-\frac{1}{\sigma_n^2} \|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2 \right). \tag{6.12}$$

Note that we do not give a mathematical derivation of (6.11) based on the derivative of the log-likelihood function. We just use that a typical decoder delivers reliability information on the transmitted sequence, which can be mapped to $p\left(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I, \mathbf{h}\right)$. The assumption that decoding is based on $p(\mathbf{y}|\mathbf{x}, \mathbf{h})$ is motivated by the log-likelihood function (6.5).

Now, based on the following set of equations, the optimization problem in (6.4) is tried to be solved iteratively

$$\hat{\mathbf{h}}^{(n)} = \left\{ \sum_{\mathbf{x}_i \in \mathcal{P}_x} p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right) \frac{1}{\sigma_n^2} \mathbf{X}_i \mathbf{X}_i^H + \mathbf{R}_h^{-1} \right\}^{-1}$$

$$\times \sum_{\mathbf{x}_i \in \mathcal{P}_x} p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right) \frac{1}{\sigma_n^2} \mathbf{X}_i^* \mathbf{y} \qquad (6.13)$$

$$\left\{ p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n)}, \hat{\mathbf{h}}^{(n)}\right) \right\} = \text{Dec}\left(p(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}^{(n)}), \forall \mathbf{x}_i \in \mathcal{P}_\mathbf{x} \right). \qquad (6.14)$$

Here the decoder, identified by the function *Dec*, calculates the soft-information on the transmitted sequence $\mathbf{x}$ based on the channel estimate $\hat{\mathbf{h}}^{(n)}$. This soft-information is used to determine the probabilities of the sequences $\mathbf{x}_i$, i.e., $p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n)}, \hat{\mathbf{h}}^{(n)}\right)$. The superscript (n) is the iteration number. Thus, following the principle of synchronized detection, the detector/decoder here uses the channel estimate as it would be the actual channel fading weight, which is indicated by writing $p(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}^{(n)})$ in (6.14). In [98] it has been shown that in case of convergence to the correct fixed point, the solution of the equation set (6.13) and (6.14) corresponds to the solution of the joint optimization problem in (6.4) and thus of (6.3). However, in [98] no statements on the convergence requirements of the iterative process are given.

Note, the approach to find a solution for the joint ML-detection and MAP parameter estimation problem in (6.3) based on iteratively solving a set of fixed point equations, corresponds to the solution of an optimization problem, i.e., the search for a point in a multidimensional space. In contrast, the question on the achievable rate, as it is the main topic of the present work, is related to the convergence region. In this regard consider that the achievable rate is related to the amount of different codewords, normalized to the codeword length, that can be distinguished by the receiver in the limit of an infinite codeword length. On the other hand, the hereby determined packing of the codewords is related to the convergence region of the iterative solution of the set of fixed point equations described before.

As it has already been stated, the aim of the work given in [98] is to show that the joint optimization problem in (6.3) corresponds to a specific fixed point solution of the equation system given in (6.13) and (6.14). For details we refer to [98, Chapter 7]. It is important to recognize that the components described by the equations in (6.13) and (6.14), i.e., the channel estimator and the decoder, are not optimal in general. The only statement that can be given is that the correct fixed point solution of (6.13) and (6.14) corresponds to the solution of the optimization problem in (6.3). That means that only in the neighborhood of the correct fixed point, the channel estimator becomes optimal in the sense that it exploits all available information. To see this, we assume that in the first iteration there is no reliability information on the transmit symbol x_k. This leads to the fact that the corresponding soft-symbol

$\tilde{x}_k$ is zero[2]. Thus, y_k is not used for channel estimation. Nevertheless y_k also contains information on the channel, which thus is not exploited. Obviously, (6.13) is in general not a MAP estimator, as it does not exploit all available information. However, at the correct fixed point, corresponding to perfect knowledge on the transmitted sequence $\mathbf{x}$, the channel estimator in (6.13) uses all available information.

The approach to express the joint optimization problem in (6.3) based on a set of fixed point equations results in decoding based on using the channel estimate as it would be the actual channel realization. We have already argued in Section 5.1.2 that using the channel estimate as it would be the actual channel realization leads to mismatched decoding, as the channel estimation is characterized by a channel estimation error. Therefore, corresponding to the discussion in Section 5.1.2, let us for the moment assume that the metric of the decoder is based on $p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}})$, i.e., accounts for the channel estimation error, and therefore, decoding is expressed by

$$\left\{ p\left(\mathbf{x}_i \big| \mathbf{y}, \boldsymbol{\lambda}_I^{(n)}, \hat{\mathbf{h}}^{(n)} \right) \right\} = \mathrm{Dec}\left(p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}^{(n)}), \forall \mathbf{x}_i \in \mathcal{P}_{\mathbf{x}} \right). \tag{6.15}$$

For detection/decoding the conditional PDF $p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}^{(n)})$ is required, cf. (5.24)

$$p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}^{(n)}) = \frac{\exp\left\{ -\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}^{(n)}\right)^H \left(\mathbf{X}\mathbf{R}_e^{(n)}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right)^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}^{(n)}\right) \right\}}{\pi^N \det\left(\mathbf{X}\mathbf{R}_e^{(n)}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N \right)}. \tag{6.16}$$

In contrast to the case of a purely pilot based channel estimation as considered in Chapter 5, now the channel estimation error covariance matrix $\mathbf{R}_e^{(n)}$ depends on the soft-symbols and, thus, on the iteration (n). Here, the index pil is omitted, as the channel estimation now is not only based on pilot symbols as in Chapter 5 but additionally on the data symbols. The estimation error variance at different symbol time instances, i.e., the diagonal elements of $\mathbf{R}_e^{(n)}$ are in general not equal. Furthermore, as in the case of pilot based synchronized detection the estimation error is not white, i.e., $\mathbf{R}_e^{(n)}$ is not diagonal.

As already argued in Section 5.1.2, typically decoders do not exploit the temporal correlation of the channel estimation error. In addition, due to interleaving it is spread over the whole length N of the observation sequence. Neglecting the temporal correlation of the estimation error leads to the approximate PDF

---

[2] We assume here zero-mean input symbols, see Section 6.2.1.

$$p(\mathbf{y}|\mathbf{x},\hat{\mathbf{h}}^{(n)}) \approx \frac{1}{\pi^N \prod_{k=1}^{N} \left(|x_k|^2 \sigma_{e_k}^{2,(n)} + \sigma_n^2\right)} \exp\left\{-\sum_{k=1}^{N} \frac{\left|y_k - x_k \hat{h}_k^{(n)}\right|^2}{|x_k|^2 \sigma_{e_k}^{2,(n)} + \sigma_n^2}\right\}$$

$$(6.17)$$

where $\sigma_{e_k}^{2,(n)}$ are the diagonal entries of $\mathbf{R}_e^{(n)}$. Notice, decoding with respect to this metric then is related to synchronized detection corresponding to Definition 5.1 stating that the estimation error is treated by the detector/decoder as additive white noise. As receivers based on iterative code-aided synchronized detection use a symbol-wise detection metric (coherent detection), the following derivation of the upper bound on the achievable rate with such a receiver structure is based on the assumption that detection/decoding uses the metric corresponding to (6.17), i.e., using (6.15) with (6.17).

In addition, consider the following. Also the effect that the noise variance $|x_k|^2 \sigma_{e_k}^{2,(n)} + \sigma_n^2$ depends on the individual time instant k is typically not used in decoding, leading to the following further approximation of $p(\mathbf{y}|\mathbf{x},\hat{\mathbf{h}}^{(n)})$, cf.(5.26)

$$p(\mathbf{y}|\mathbf{x},\hat{\mathbf{h}}^{(n)}) \approx \frac{1}{\pi^N (\sigma_x^2 \sigma_e^2 + \sigma_n^2)^N} \exp\left\{-\frac{\|\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}^{(n)}\|^2}{\sigma_x^2 \sigma_e^2 + \sigma_n^2}\right\} \qquad (6.18)$$

where we have approximated $|x_k|^2 \sigma_{e_k}^{2,(n)}$ by $\sigma_x^2 \sigma_e^2$. Typically the estimation error variances $\sigma_{e_k}^{2,(n)}$ are unknown to the receiver, except for the case that the channel estimation is solely based on pilot symbols. In this case, it can be calculated analytically. Therefore, suitable assumptions for σ_e^2 in (6.18) have to be made. In [40] the effect of a mismatch between the actual estimation error variance and the value used for decoding has been examined based on Monte Carlo simulations. This examination shows, that the performance in terms of the BER of the iterative synchronization and decoding algorithm is robust in case the estimation error variance assumed for decoding is smaller than the actual value. Therefore, one reasonable choice of σ_e^2 is to assume it corresponding to the case of perfect decoder feedback, i.e., calculating the channel estimation error variance under the assumption that all data symbols are perfectly known. However, the approximation given with (6.18) is not considered for the calculation of the upper bound on the achievable rate in the following.

6.1.1 Modified Channel Estimation Unit

As it has already been stated, we have not been able to derive an upper bound on the achievable rate with the receiver described by (6.13) and (6.15) with (6.17), but only for a slightly modified version. The modified receiver has a

slightly different channel estimator, which for the calculation of the estimate $\hat{h}_k$, i.e., the k-th entry of $\hat{\mathbf{h}}$, does not use the corresponding observation y_k. Therefore, the channel estimator in (6.13) is substituted by the following one:

$$
\hat{h}_k^{(n)} = \left[\left\{ \sum_{\mathbf{x}_i \in \mathcal{P}_x} p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right) \frac{\mathbf{B}_k \mathbf{X}_i \mathbf{X}_i^H}{\sigma_n^2} + \mathbf{R}_h^{-1} \right\}^{-1} \right.
$$
$$
\left. \times \sum_{\mathbf{x}_i \in \mathcal{P}_x} p\left(\mathbf{x}_i | \mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right) \frac{\mathbf{B}_k \mathbf{X}_i^*}{\sigma_n^2} \mathbf{y} \right]_k, \forall k \qquad (6.19)
$$

where $[\mathbf{a}]_k$ denotes the k-th element of the vector $\mathbf{a}$. Furthermore, $\mathbf{B}_k$ is a diagonal matrix, whose diagonal elements are given by

$$
[\mathbf{B}_k]_{l,l} = \begin{cases} 1 \text{ for } l \neq k \\ 0 \text{ for } l = k \end{cases}. \qquad (6.20)
$$

This means that by multiplication of $\mathbf{X}_i$ with $\mathbf{B}_k$ the k-th diagonal element of $\mathbf{X}_i$ is multiplied by zero, corresponding to the case of not using the observation y_k and the soft-information on x_k for the estimation of $\hat{h}_k$. Although it is not the typically studied channel estimator for code-aided channel estimation, this channel estimator is also known from literature, see, e.g., [95].

6.2 Achievable Rate with Iterative Code-Aided Synchronized Detection

Before calculating bounds on the achievable rate with a receiver following the principle of iterative code-aided synchronized detection as it has been presented in Section 6.1, we want to get an understanding, which gain in principle can be achieved while iteratively enhancing the channel estimation based on soft-information on the data symbols delivered by the decoder. Therefore, recall the following expression for $\mathcal{I}(\mathbf{y}; \mathbf{x})$ derived in (5.29)

$$
\mathcal{I}(\mathbf{y}; \mathbf{x}) = \mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D}\right). \qquad (6.21)
$$

Obviously, with iterative code-aided synchronized detection we are not able to get a mutual information larger than $\mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D}\right)$. Now, recall that in case of synchronized detection with a solely pilot based channel estimation as it has been studied in Chapter 5, only parts of $\mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\mathrm{pil},D}\right)$ can be used. The reason for this is that the interpolation based channel estimation leads to an estimation error process $\{e_{\mathrm{pil},k}\}$ with $e_{\mathrm{pil},k}$ defined in (5.9) which is temporally correlated [90]. The detector/decoder does not exploit the temporal correlation of the channel estimation error process. Therefore, in

Section 5.2 we have introduced the mutual information $\mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right)$, where the artificial channel estimation vector $\check{\hat{\mathbf{h}}}_{\mathrm{pil},D}$ corresponds to $\hat{\mathbf{h}}_{\mathrm{pil},D}$, except that the temporal correlation of the corresponding channel estimation error process is assumed to be white. Hence, $\mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right)$ is smaller than $\mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D \left| \hat{\mathbf{h}}_{\mathrm{pil},D} \right.\right)$, see (5.31). Thus, the following difference is an upper bound on the possible gain while using iterative code-aided synchronized detection instead of synchronized detection with a solely pilot based channel estimation

$$G = \mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D \left| \hat{\mathbf{h}}_{\mathrm{pil},D} \right.\right) - \mathcal{I}\left(\mathbf{y}_D; \mathbf{x}_D \left| \check{\hat{\mathbf{h}}}_{\mathrm{pil},D} \right.\right). \tag{6.22}$$

This means that in case of convergence of the iterative synchronized detection based receiver to the case of error-free decoding, the receiver based on iterative synchronization and decoding can retrieve parts of the information that have been discarded by ignoring the temporal correlation of the channel estimation error process in case of synchronized detection in combination with a solely pilot based channel estimation. Thus, for a given pilot spacing the difference G gives an upper bound on the possible gain by iterative code-aided synchronized detection in comparison to solely pilot based synchronized detection. The difference G will depend on the pilot spacing. However, it is not assured if the iterative receiver will converge to error-free decoding. Furthermore, we have not shown that the first term on the RHS of (6.22), i.e., $\mathcal{I}\left(\mathbf{x}_D; \mathbf{y}_D \left| \hat{\mathbf{h}}_{\mathrm{pil},D} \right.\right)$ can be achieved by iterative code-aided synchronized detection. Therefore, the difference G is only an upper bound to the maximum possible gain while using iterative synchronization and decoding in comparison to solely pilot based synchronized detection.

6.2.1 Upper-Bounding Approach on the Achievable Rate

One of the main goals of the present chapter is the derivation of an upper bound on the achievable rate with a receiver following the principle of synchronized detection in combination with a code-aided channel estimation as it has been discussed previously, i.e., a receiver that is described by the modified channel estimator in (6.19) and by a detector using a symbol-wise metric, i.e., described by using (6.15) in combination with (6.17).

Before starting with the derivation of the upper bound on the achievable rate with such a receiver structure, we shortly describe the approach we will take to calculate this upper bound. Let us for the moment assume that the transmitted sequence consists out of i.i.d. data symbol and no pilot symbols. Now, bear in mind that we use a detector using a symbol-wise detection

metric, see (6.17). I.e., the detector is not able to exploit any information contained in the temporal correlation of the channel estimation error. This allows us to evaluate the achievable rate based on the mutual information at an arbitrarily chosen time instant k, cf. the discussion in Section 5.2. Furthermore, consider that due to our assumption on i.i.d. transmit symbols the only dependency between the individual time instances is established by the channel correlation. To evaluate the achievable rate at the time instant k, we have to consider the knowledge on h_k given by all time instances different from k. I.e., we want to express the information on the channel given by all time instances different from the arbitrarily chosen time instant k by an estimate $\hat{h}_k$ of the channel at the time instant k. Now, we want to calculate an upper bound on the mutual information between the channel input x_k and the channel output y_k. Obviously, the mutual information between x_k and y_k is maximized if the estimation error variance of the channel estimate $\hat{h}_k$ is minimized. More frankly speaking, the better the channel estimate, i.e., the smaller the channel estimation error variance, the larger is the mutual information between x_k and y_k. To get an upper bound on the achievable rate, we assume that all transmit symbols in the past and in the future are known, which obviously lower-bounds the channel estimation error variance. I.e., an upper bound on the achievable rate at the time instant k is given by the mutual information $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$. Note that the estimation error variance of $\hat{h}_k$ depends on the distribution of all transmit symbols except the one at the time instant k, i.e., on $\mathbf{x}_{\backslash k}$. Here $\mathbf{x}_{\backslash k}$ corresponds to $\mathbf{x}$ without the element x_k. This is the reason, why it has to be additionally conditioned on $\mathbf{x}_{\backslash k}$ and not only on $\hat{h}_k$. I.e., $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$ expresses the mutual information between the channel input and the channel output at the time instant k, when *knowing* the channel estimate $\hat{h}_k$. In the following we will show that this estimate $\hat{h}_k$ is an MMSE estimate of h_k based on all channel output observations except of y_k, i.e., $\mathbf{y}_{\backslash k}$ and all transmit symbols except of x_k, i.e., $\mathbf{x}_{\backslash k}$. Due to the Gaussian nature of the problem the estimator is linear. Thus, the channel estimate $\hat{h}_k$ exactly corresponds to the channel estimate given by the modified channel estimator in (6.19) if all past and all future transmit symbols are perfectly known. Note that the MMSE estimate $\hat{h}_k$ is only based on $\mathbf{y}_{\backslash k}$ and $\mathbf{x}_{\backslash k}$, i.e., it does not use y_k, like the channel estimator in (6.19). I.e., the estimator $\hat{h}_k$ corresponds to an interpolation.

In conclusion, $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$ is an upper bound on the achievable rate when using a receiver based on iterative code-aided channel estimation, as it is described by the channel estimator in (6.19), which does not use y_k for the estimation of h_k and the coherent detector described by (6.15) in combination with (6.17). In the following section this statement will be formally derived.

Note that this statement holds also in case we use pilot symbols, as long as we suppose that the time instant k is used to transmit a data symbols and not a pilot symbol. This should be obvious, as the term $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$, which is used as an upper bound on the achievable rate with the given re-

ceiver structure, corresponds to the case that all past and all future transmit symbols are known. At this point it is irrelevant if these symbols are known data symbols or pilot symbols.

The mutual information $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$ depends on the estimation error variance of $\hat{h}_k$, which itself is a random quantity, whose distribution depends on the distribution of the past and future transmit symbols contained in $\mathbf{x}_{\backslash k}$. This is analogous to the case of channel prediction discussed in Chapter 4. In the case of channel prediction it has been shown that an upper bound on the achievable rate is given if the channel prediction error variance is calculated under the assumption that all symbols in the past are constant modulus symbols. In contrast, we have no proof that the mutual information rate corresponding to $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$, i.e., considering an infinitely long past and future for the calculation of the estimation error variance, is also maximized by the calculation of the channel interpolation error variance under the assumption that all past and future transmit symbols with respect to the time instant k are constant modulus symbols. Therefore, we can only give a non-closed form solution for an upper bound on the achievable rate, which still depends on the channel interpolation error variance and, thus, on the distribution of all past and future transmit symbols. However, for small channel dynamics it seems reasonable that the channel interpolation error variance can be well approximated by the interpolation error variance calculated under the assumption that all past and future transmit symbols are constant modulus symbols. Notice that the assumption on constant modulus input symbols is only made in the context of the interpolation error variance and not to calculate the mutual information at the time instant k. Therefore, in the latter context x_k is still assumed to be arbitrarily distributed.

In the derivation below we use the following assumptions on the data symbols. We assume i.i.d. zero-mean input symbols. The motivation for the fact that the symbols have zero-mean is given on the one hand by the fact that in the coherent case the capacity-achieving input distribution also is i.i.d. zero-mean, and on the other hand by the fact that in case of pilot based synchronized detection the rate maximizing data symbols are also zero-mean. The assumption on i.i.d. input symbols significantly simplifies the presentation of the derivation, and, as it has already been discussed before, still allows the incorporation of pilot symbols in the final step of the calculation of the upper bound.

Corresponding to Chapter 5, we do not use any peak power constraint so that it is in general optimal to use the maximum average transmit power, i.e., $\mathrm{E}\left[|x_k|^2\right] = \sigma_x^2$, cf. (4.1). Hence, the parameter ρ defined in (2.18) corresponds to the average SNR.

6.2.2 The Channel Interpolation Separation Inequality

Motivated by the preceding discussion, within the present section we derive an upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ based on a separation of the transmission in past time instances, the present time instant, and future time instances. Hereby, we separate the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ into the term $\mathcal{I}(y_k; x_k | \hat{h}_k, \mathbf{x}_{\backslash k})$, which, as motivated before is shown to be an upper bound to the achievable rate with the given receiver structure, and an additional term, which cannot be exploited by the receiver described by (6.19) and (6.15) in combination with (6.17). In the following, we refer to such a separation as *channel interpolation separation*.

As already known from (3.2), the mutual information rate between the input and the output of the channel is defined by

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \lim_{N \to \infty} \frac{1}{2N+1} \mathcal{I}(\mathbf{y}; \mathbf{x}) \tag{6.23}$$

where now the vectors $\mathbf{x}$ and $\mathbf{y}$ are of length $2N + 1$. In this chapter, we assume without loss of generality that the length of the vector $\mathbf{x}$ is odd and that its elements have the following mapping to the time instances

$$\mathbf{x} = [x_{-N}, \dots, x_{-1}, x_0, x_1, \dots, x_N]^T. \tag{6.24}$$

The vectors $\mathbf{y}$ and $\mathbf{h}$ are constructed correspondingly.

Now, we separate the mutual information $\mathcal{I}(\mathbf{y}; \mathbf{x})$ in a way that will allow us to bound the achievable rate with iterative code-aided synchronized detection as described in Section 6.2.1. Therefore, we introduce the vector

$$\mathbf{x}_{\backslash 0} = [x_{-N}, \dots, x_{-1}, x_1, \dots, x_N]^T \tag{6.25}$$

i.e., the element at the time instant 0 is discarded. In the following, we use a corresponding notation for $\mathbf{y}_{\backslash 0}$. Here the time instant 0 corresponds to the arbitrary time instant k in the description of the bounding approach given in Section 6.2.1.

To derive an upper bound on $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ based on the channel interpolation separation, we separate $\mathcal{I}(\mathbf{y}; \mathbf{x})$ as follows

$$
\begin{aligned}
\mathcal{I}(\mathbf{y}; \mathbf{x}) &\overset{(a)}{=} \mathcal{I}(y_0; \mathbf{x} | \mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{x}; \mathbf{y}_{\backslash 0}) \\
&\overset{(b)}{=} \mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{x}; \mathbf{y}_{\backslash 0}) \\
&\overset{(c)}{=} \mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{x}_{\backslash 0}; \mathbf{y}_{\backslash 0} | x_0) + \mathcal{I}(x_0; \mathbf{y}_{\backslash 0}) \\
&\overset{(d)}{=} \mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{x}_{\backslash 0}; \mathbf{y}_{\backslash 0}) \tag{6.26}
\end{aligned}
$$

where for (a)-(c), we have used the chain rule for mutual information. Finally, for (d) we have used the independency of the transmit symbols, and therefore

$$\mathcal{I}(x_0; \mathbf{y}_{\backslash 0}) = 0 \tag{6.27}$$

and

$$\mathcal{I}(\mathbf{x}_{\backslash 0}; \mathbf{y}_{\backslash 0}|x_0) = \mathcal{I}(\mathbf{x}_{\backslash 0}; \mathbf{y}_{\backslash 0}). \tag{6.28}$$

Obviously (6.26) is equivalent to

$$\mathcal{I}(\mathbf{y}; \mathbf{x}) - \mathcal{I}(\mathbf{x}_{\backslash 0}; \mathbf{y}_{\backslash 0}) = \mathcal{I}(y_0; x_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}). \tag{6.29}$$

In the following, we show that for $N \to \infty$ the LHS of (6.29) is an upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$. Therefore, we rewrite the LHS of (6.29) as follows

$$
\begin{aligned}
\mathcal{I}(\mathbf{y}; \mathbf{x}) - \mathcal{I}(\mathbf{y}_{\backslash 0}; \mathbf{x}_{\backslash 0}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) - h(\mathbf{y}_{\backslash 0}) + h(\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}) \\
&\overset{(a)}{=} h(y_0|\mathbf{y}_{\backslash 0}) + h(\mathbf{y}_{\backslash 0}) - h(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x}) - h(\mathbf{y}_{\backslash 0}|\mathbf{x}) - h(\mathbf{y}_{\backslash 0}) + h(\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}) \\
&\overset{(b)}{=} h(y_0|\mathbf{y}_{\backslash 0}) - h(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x}) \\
&= \mathcal{I}(y_0; \mathbf{x}|\mathbf{y}_{\backslash 0}) \\
&\overset{(c)}{\geq} \mathcal{I}(y_0; \mathbf{x}|\mathbf{y}_{-N}^{-1}) \\
&\overset{(d)}{=} \mathcal{I}(y_0; \mathbf{x}_{-N}^0|\mathbf{y}_{-N}^{-1}) \\
&= h(y_0|\mathbf{y}_{-N}^{-1}) - h(y_0|\mathbf{x}_{-N}^0, \mathbf{y}_{-N}^{-1}) \tag{6.30}
\end{aligned}
$$

where, e.g., $\mathbf{y}_{-N}^{-1}$ is a subvector of $\mathbf{y}$ containing the symbols from time instant $-N$ to -1. Equality (a) is based on the chain rule for differential entropy, and for (b) we have used that

$$h(\mathbf{y}_{\backslash 0}|\mathbf{x}) = h(\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}) \tag{6.31}$$

due to the independency of the transmit symbols. Furthermore, inequality (c) holds as the knowledge on the channel output observations $\mathbf{y}_1^N$ will increase the mutual information between y_0 and $\mathbf{x}$ as they contain information on h_0. Additionally, (d) holds due to the independency of the transmit symbols and the fact that knowledge of the future transmit symbols $\mathbf{x}_1^N$ does not help to estimate the channel weight h_0 in case the channel outputs $\mathbf{y}_1^N$ are not known.

Taking finally the limit $N \to \infty$ of the RHS of (6.30) yields

$$\lim_{N \to \infty} \left\{ h(y_0|\mathbf{y}_{-N}^{-1}) - h(y_0|\mathbf{x}_{-N}^0, \mathbf{y}_{-N}^{-1}) \right\} = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}) \tag{6.32}$$

using the definition of the entropy rate for stationary ergodic processes [19, Chapter 4.2].

Hence, with (6.29), (6.30), and (6.32) we get the following upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})^3$,

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \lim_{N \to \infty} \left\{ \mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0}) \right\}. \qquad (6.34)$$

In the following, we discuss the two terms at the RHS of (6.34). We argue that the second term at the RHS of (6.34), i.e., $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0})$ cannot be exploited by the receiver structure described by (6.19) and (6.15) in combination with (6.17). In this regard note that $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0})$ is a component of the upper bound on the mutual information rate between the transmitter and the receiver. However, a corresponding term also arises when using the prediction separation in (6.33), i.e., $\mathcal{I}(y_N; \mathbf{x}_1^{N-1} | \mathbf{y}_1^{N-1})$, which does not yield an upper bound, but the actual mutual information rate. As the second term at the RHS of (6.34) cannot be exploited by the discussed receiver structure, by calculating an upper bound on the first term on the RHS of (6.34), i.e., $\mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$, we will get an upper bound on the achievable rate with the iterative code-aided synchronized detection based receiver structure described by (6.19) and (6.15) in combination with (6.17). We start with the discussion of the first term, i.e., $\mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$.

6.2.3 The Term $\mathcal{I}(y_0; x_0 | \mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$

Obviously, $\mathcal{I}(y_0; x_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0})$ corresponds to the mutual information between y_0 and x_0 when knowing the transmit symbols and the channel output at all time instances except of zero. This exactly corresponds to the assumptions we discussed in Section 6.2.1 in the context of describing the approach for the derivation of an upper bound on the achievable rate with the receiver based on iterative synchronization and decoding described by (6.19) and (6.15) in combination with (6.17). In the following, this relation will become even more obvious by identifying that conditioning on $\mathbf{y}_{\backslash 0}$ and $\mathbf{x}_{\backslash 0}$ is equivalent to conditioning on the MMSE estimate $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$ which is based on $\mathbf{y}_{\backslash 0}$ and $\mathbf{x}_{\backslash 0}$.

Similar to the argumentation in the context of the channel predictor in Section 4.1, conditioning on $\mathbf{x}_{\backslash 0}$ and $\mathbf{y}_{\backslash 0}$ is equivalent to the conditioning

---

^{3} Notice a similar derivation based on a separation of the transmit interval into the current time instant N and all previous time instances $1, \ldots, N-1$ leads to the following equality

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \lim_{N \to \infty} \left\{ \mathcal{I}(y_N; x_N | \mathbf{y}_1^{N-1}, \mathbf{x}_1^{N-1}) + \mathcal{I}(y_N; \mathbf{x}_1^{N-1} | \mathbf{y}_1^{N-1}) \right\}. \qquad (6.33)$$

In case of such a prediction separation we get an equality while in case of the interpolation separation in (6.34) we get an inequality. This difference arises due to the fact that using the chain rule of mutual information and the definition of the mutual information rate for a stationary stochastic process, $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ can be canonically expressed based on channel prediction while this is not possible for interpolation.

on the MMSE interpolation $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$. This results from the fact that $\hat{h}_0$ contains all information on h_0 being contained in $\mathbf{y}_{\backslash 0}$ while knowing $\mathbf{x}_{\backslash 0}$, which will be shown in the following. As $\hat{h}_0$ is an MMSE estimate, it is given by

$$\hat{h}_0 = \mathrm{E}\left[h_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right]. \tag{6.35}$$

Analogous to the prediction case, it can be shown that the estimate $\hat{h}_0$ is zero-mean proper Gaussian. Due to the fact that the estimation error

$$e_0 = h_0 - \hat{h}_0 \tag{6.36}$$

and the estimate $\hat{h}_0$ are independent, it follows that the estimation error e_0 is also zero-mean proper Gaussian. Its variance is given by

$$\begin{aligned}
\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}) &= \mathrm{E}\left[\left|h_0 - \hat{h}_0\right|^2 \Big| \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] \\
&= \mathrm{E}\left[|e_0|^2 \Big| \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] \\
&\overset{(a)}{=} \mathrm{E}\left[|e_0|^2 \Big| \mathbf{x}_{\backslash 0}\right]
\end{aligned} \tag{6.37}$$

where the index *int* denotes interpolation. Equality (a) is based on the fact that the estimation error is orthogonal to and, thus, independent of the observation vector $\mathbf{y}_{\backslash 0}$. Obviously, the interpolation error variance depends on the past and future transmit symbols, which is indicated by the notation $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$. Therefore, the variance of the estimate $\hat{h}_0$ is given by

$$\sigma_{\hat{h}}^2(\mathbf{x}_{\backslash 0}) = \sigma_h^2 - \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}). \tag{6.38}$$

Following the same argumentation as in case of the channel predictor, see Section 4.1, the channel output y_0 conditioned on $\mathbf{y}_{\backslash 0}, \mathbf{x}$ is proper Gaussian with mean, cf. (4.5)

$$\begin{aligned}
\mathrm{E}\left[y_0 | \mathbf{x}, \mathbf{y}_{\backslash 0}\right] &= \mathrm{E}\left[x_0 h_0 + n_0 | \mathbf{x}, \mathbf{y}_{\backslash 0}\right] \\
&= x_0 \mathrm{E}\left[h_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] \\
&= x_0 \hat{h}_0.
\end{aligned} \tag{6.39}$$

and variance

$$\text{var}\left[y_0|\mathbf{x}, \mathbf{y}_{\backslash 0}\right] = \text{E}\left[\left|y_0 - \text{E}\left[y_0|\mathbf{x}, \mathbf{y}_{\backslash 0}\right]\right|^2 \Big| \mathbf{x}, \mathbf{y}_{\backslash 0}\right]$$

$$= \text{E}\left[\left|y_0 - x_0\hat{h}_0\right|^2 \Big| \mathbf{x}, \mathbf{y}_{\backslash 0}\right]$$

$$= |x_0|^2\text{E}\left[\left|h_0 - \hat{h}_0\right|^2 \Big| \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] + \sigma_n^2$$

$$= |x_0|^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_n^2. \tag{6.40}$$

Thus, the PDF $p(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x})$ is given by

$$p(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x}) = \frac{1}{\pi(|x_0|^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_n^2)} \exp\left(-\frac{|y_0 - \hat{h}_0 x_0|^2}{|x_0|^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_n^2}\right)$$

$$= p(y_0|\hat{h}_0, \mathbf{x}). \tag{6.41}$$

For comparison see (4.12) in the context of the channel prediction. Using (6.41) also $p(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$ can be expressed by

$$p(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0}) = \int p(y_0|\mathbf{y}_{\backslash 0}, \mathbf{x})p(x_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})dx_0$$

$$= \int p(y_0|\hat{h}_0, \mathbf{x})p(x_0)dx_0$$

$$= p(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}). \tag{6.42}$$

Based on (6.41) and (6.42) the following identity holds

$$\mathcal{I}(x_0; y_0|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}) = \mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}). \tag{6.43}$$

Thus, the first term on the RHS of (6.34) is the mutual information between the input and the output of the channel if an MMSE estimate $\hat{h}_0$ is available, which is based on all past and future channel observations and the knowledge of all transmit symbols except the current one.

We want to explain why $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ is an upper bound on the achievable rate with the iterative code-aided channel estimation based receiver given by (6.19) and (6.15) in combination with (6.17). In this regard, consider that the only dependency between the individual time instances is established by the channel correlation. In case all past and all future transmit symbols are known, corresponding to $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$, all information on h_0 given by the past and future channel output observation $\mathbf{y}_{\backslash 0}$ and the knowledge on $\mathbf{x}_{\backslash 0}$ is carried by $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$. Note that for the calculation of $\hat{h}_0$, the observation y_0 is not used, which exactly corresponds to the channel estimator described by (6.19). Furthermore, observe that for the case of perfect knowledge of all past and all future transmit symbols the estimator in (6.19) exactly corresponds to the MMSE interpolator $\hat{h}_0$ in $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ with the estimation error variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$ given in (6.37). This means that the term $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$

corresponds to the mutual information at the arbitrarily chosen data symbol time instant 0 if all past and all future transmit symbols are perfectly known. Obviously, the assumption of perfect knowledge of all past and all future transmit symbols results in an upper bound to the actual achievable mutual information at the arbitrarily chosen data symbol time instant 0.

Furthermore, it is important to recognize that this argumentation only holds in case we assume a symbol-wise detection as described by (6.17). If detection would be performed over the whole sequence, evaluation of the mutual information for a single time instant, as it is done with $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ would be meaningless, as the information contained in the temporal correlation of the channel estimation error is not captured. In this context see also the discussion in Section 5.2.

Based on this argumentation, we are now able to interpret the representation of the mutual information rate in (6.34). It is the sum of two terms. The first term on the RHS of (6.34), which is the main contribution, is related to a *coherent Rayleigh flat-fading channel*, i.e., a channel whose optimum detection metric can be evaluated symbol-wise. In comparison to the genuine fading channel its fading variance is modified due to the estimation error and given by $\sigma_h^2 - \sigma_e^2$. In addition, its noise variance is given by $|x_k|^2 \sigma_e^2 + \sigma_n^2$. However, the fact that the *effective* noise variance depends on the transmit symbol x_k is a difference to a coherent fading channel. Furthermore, the second term on the RHS of (6.34) can be viewed as a correction term which describes the amount of information contained in the temporal correlation of the channel estimation error, which cannot be exploited by the symbol-wise, i.e., coherent, detector. Thus, the first term on the RHS of (6.34) accounts for the detector imperfection. This imperfection is corrected by the second term, which is then dependent on the correlation of the channel estimation error. However, the second term cannot be exploited by the receiver structure described by (6.19) and (6.15) in combination with (6.17).

6.2.4 The Term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$

As already discussed, the information contained in the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ cannot be exploited by the receiver structure described by (6.19), (6.15), and (6.17). In the present section, we try to interpret the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ in more detail. Therefor, consider that the only relation between the individual symbol time instances is established by the correlation of the channel fading process. Thus, using the chain rule for mutual information, we rewrite $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ such that its relation to the fading weight h_0 becomes explicit:

$$\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) = \mathcal{I}(h_0, y_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) - \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y})$$
$$= \mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|h_0, \mathbf{y}_{\setminus 0}) + \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) - \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y})$$
$$\overset{(a)}{=} \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) - \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y}) \tag{6.44}$$

where for (a) we have used that $\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|h_0, \mathbf{y}_{\setminus 0})$ is zero, as

$$\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|h_0, \mathbf{y}_{\setminus 0}) = h(y_0|h_0, \mathbf{y}_{\setminus 0}) - h(y_0|\mathbf{x}_{\setminus 0}, h_0, \mathbf{y}_{\setminus 0})$$
$$\overset{(b)}{=} h(y_0|h_0) - h(y_0|h_0) = 0 \tag{6.45}$$

where (b) follows from the fact that y_0 is independent of $\mathbf{y}_{\setminus 0}$ and of $\mathbf{x}_{\setminus 0}$ while conditioning on h_0.

Expressing the RHS of (6.44) in terms of differential entropies allows to discuss the meaning of the term $\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0})$ more easily:

$$\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) = \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) - \mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y})$$
$$= \big(h(h_0|\mathbf{y}_{\setminus 0}) - h(h_0|\mathbf{y}_{\setminus 0}, \mathbf{x}_{\setminus 0})\big) - \big(h(h_0|\mathbf{y}) - h(h_0|\mathbf{y}, \mathbf{x}_{\setminus 0})\big)$$
$$\overset{(a)}{=} \Big(h(h_0|\mathbf{y}_{\setminus 0}) - h(h_0|\hat{h}_0, \mathbf{x}_{\setminus 0})\Big) - \Big(h(h_0|\mathbf{y}) - h(h_0|\hat{h}_0, y_0, \mathbf{x}_{\setminus 0})\Big)$$
$$\tag{6.46}$$

where for (a) we have substituted the condition on $\mathbf{y}_{\setminus 0}, \mathbf{x}_{\setminus 0}$ by the condition on $\hat{h}_0, \mathbf{x}_{\setminus 0}$, where $\hat{h}_0$ is the MMSE estimate of h_0, which is based on the knowledge of $\mathbf{y}_{\setminus 0}$ and $\mathbf{x}_{\setminus 0}$, cf. (6.35), see Section 6.2.3.

The difference on the RHS of (6.46) is related to the amount of information on the channel fading weight h_0 that can be retrieved, on the one hand, while using the observation y_0, and on the other hand, while not using it. Let us first look at the difference in the first pair of brackets on the RHS of (6.46), i.e.,

$$\mathcal{I}(h_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0}) = \Big(h(h_0|\mathbf{y}_{\setminus 0}) - h(h_0|\hat{h}_0, \mathbf{x}_{\setminus 0})\Big). \tag{6.47}$$

The first term on the RHS of (6.47) is the uncertainty in h_0 while conditioning on the observation sequence $\mathbf{y}_{\setminus 0}$. Furthermore, the second term, i.e., $h(h_0|\hat{h}_0, \mathbf{x}_{\setminus 0})$ is the remaining uncertainty in h_0, when we condition on the MMSE estimate $\hat{h}_0$ and on $\mathbf{x}_{\setminus 0}$. Thus, this term corresponds to the entropy of the channel estimation error. Overall, the difference on the RHS of (6.47) is related to the reduction of uncertainty in h_0 while *knowing* the observation sequence $\mathbf{y}_{\setminus 0}$, when using the channel estimator which exploits the knowledge on $\mathbf{x}_{\setminus 0}$.

Analogously, the difference in the second pair of brackets on the RHS of (6.46), i.e.,

$$\mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y}) = \left(h(h_0|\mathbf{y}) - h(h_0|\hat{h}_0, y_0, \mathbf{x}_{\backslash 0}) \right) \tag{6.48}$$

corresponds to the difference in (6.47) discussed before, except that in all terms additionally it is conditioned on y_0. I.e., the second term on the RHS of (6.48) is the remaining uncertainty in h_0 when conditioning on y_0 in addition to the MMSE estimate $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$. This entropy can be understood as the entropy of the estimation error of an enhanced channel estimator, which additionally uses the observation y_0. Also in the first term on the RHS of (6.48) it is conditioned on y_0, so that the difference in (6.48) is related to the reduction of uncertainty in h_0 while *knowing* the observation sequence $\mathbf{y}$, when using the channel estimator which exploits the knowledge on $\mathbf{x}_{\backslash 0}$.

The term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ is equal to the difference of (6.47) and (6.48). Therefore, it is related to the additional information on h_0 contained in y_0 while *knowing* $\mathbf{y}_{\backslash 0}$. The fact that this additional information on h_0 contained in y_0 cannot be exploited by the given receiver structure, i.e., using the channel estimator in (6.19) in combination with the detector in (6.15) and the symbol-wise detection metric in (6.17), is supported by the following observation. First, consider that the observation y_0 is also used at the input of the detection unit. Nevertheless, with the given structure, where y_0 is not used for channel estimation, the information contained in $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ cannot be exploited. At this point again consider that detection works symbol-wise, and that y_0 contains additional information on h_0, which can only be exploited when using it in combination with $\mathbf{y}_{\backslash 0}$ and $\mathbf{x}_{\backslash 0}$, indicated by the conditioning on $\mathbf{y}_{\backslash 0}$ in $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$. However, this is not possible for the detector due to its symbol-wise metric. This supports the statement at the end of Section 6.2.3 that the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ is a correction term to the first term on the RHS of (6.34), i.e., $\mathcal{I}(y_0; x_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$, accounting for the fact that the detector cannot exploit the temporal correlation of the channel estimation error.

This observation, i.e., that the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ accounts for the temporal correlation of the channel estimation error, which cannot be exploited by the receiver structure given by (6.19), (6.15), and (6.17) is also supported by the fact that in case of an uncorrelated channel, i.e., $\mathrm{E}[h_k h_l^*] = 0, \quad \forall k \neq l$ the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ becomes zero.

However, as we want to derive an upper bound on the achievable rate for the iterative receiver structure described by (6.19) and (6.15) with (6.17), we do not need to consider the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ any further, as it cannot be exploited by this receiver structure.

Concerning the magnitude of the information contained in $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$, at least for small channel dynamics, the additional information on h_0 delivered by y_0 is small. The reason is that the adjacent symbols, where the transmit symbols are assumed to be known, already deliver much information on h_0. Therefore, we guess that for small channel dynamics the additional information on h_0 contained in y_0 and, thus, $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ is small.

6.2.4.1 Can $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|y_{\backslash 0})$ be exploited by using the channel estimator in (6.13)?

Obviously, the modified channel estimator in (6.19) never uses the observation y_0 to estimate $\hat{h}_0$. In contrast to this, the genuine channel estimator in (6.13) uses the observation y_0, at least, if soft-information on the transmit symbols x_0 is available. However, note that in case no reliability information on x_0 is available, the corresponding soft-symbol given by the corresponding diagonal element of $\sum_{\mathbf{x}_i \in \mathcal{P}_x} p\left(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right) \frac{\mathbf{X}_i^*}{\sigma_n^2}$ is zero, and, thus, y_0 is multiplied by zero and not used for channel estimation.

Furthermore, consider that for the initial channel estimation in the receiver no reliability information on x_0 is available, as the initial channel estimation is based on pilot symbols. However, in further channel estimation iterations there is reliability information on x_0, which is delivered by the decoder. Effectively, the decoder gains this information on x_0 using the dependencies between the transmit symbols given by the code. Due to this reliability information on x_0, in later channel estimation iterations the observation y_0 is used by the estimator in (6.13). However, it is difficult to model this in the given information theoretic framework with i.i.d. data symbols.[4]

Due to this gap between the receiver exploiting code dependencies, on the one hand, and the information theoretic treatment based on i.i.d. data symbols, on the other hand, we are not able to give an answer to the question, if and which part of the information contained in the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|y_{\backslash 0})$ can be exploited when using the genuine channel estimator in (6.13), which also uses the observation y_0, in combination with the symbol-wise detector given by (6.15) and (6.17). However, note that parts of $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ arise due to the upper-bounding in (6.34).

6.2.5 An Upper Bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$

Recalling the discussion in the previous sections, it should now be obvious that an upper bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ is also an upper bound on the achievable rate with the receiver using iterative code-aided synchronized detection described by (6.19), (6.15), and (6.17) in Section 6.1. Therefore, we calculate an upper bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$.

In this regard, we express $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ based on differential entropies:

---

[4] Note that the assumption on i.i.d. input distributions is not a contradiction to the use of a codebook with codewords, which obviously means that data symbols are not independent. When speaking about i.i.d. input distributions as they are used to calculate the mutual information, the distribution for the generation of the codebook is meant. Of course, reliable communication in the sense of a diminishing probability of an erroneous decoded codeword for infinite codeword lengths requires the use of a code.

$$\mathcal{I}(x_0; y_0 | \hat{h}_0, \mathbf{x}_{\backslash 0}) = h(y_0 | \hat{h}_0, \mathbf{x}_{\backslash 0}) - h(y_0 | x_0, \hat{h}_0, \mathbf{x}_{\backslash 0}). \qquad (6.49)$$

For the upper-bounding a similar approach is applied, as it has been used to calculate the upper bound on the achievable rate with solely pilot based synchronized detection in [6]. The difference is that in [6] the channel estimate is solely based on the pilot symbols, whereas here it is based an all past and all future symbols.

6.2.5.1 Calculation of $h(y_0 | x_0, \hat{h}_0, \mathbf{x}_{\backslash 0})$

Based on $\hat{h}_0$ the channel output y_0 can be written as

$$y_0 = x_0 h_0 + n_0 = x_0 \left(\hat{h}_0 + e_0 \right) + n_0 \qquad (6.50)$$

where $\hat{h}_0$ is the MMSE estimate of h_0 based on $\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}$, see (6.35).

As the channel estimation error e_0 is zero-mean proper Gaussian, see Section 6.2.3, the channel output y_0 conditioned on $\hat{h}_0, \mathbf{x}$, i.e., $\mathbf{y}_{\backslash 0}, \mathbf{x}$, is obviously proper Gaussian, see also Section 6.2.3. Thus, the entropy $h(y_0 | x_0, \hat{h}_0, \mathbf{x}_{\backslash 0})$ is completely described by the variance of y_0 conditioned on $\mathbf{x}$ and $\hat{h}_0$, which is given in (6.40). Thus, we can rewrite the entropy $h(y_0 | x_0, \hat{h}_0, \mathbf{x}_{\backslash 0})$ as follows

$$h(y_0 | x_0, \hat{h}_0, \mathbf{x}_{\backslash 0}) = \mathrm{E}_{x_0} \left[\mathrm{E}_{\mathbf{x}_{\backslash 0}} \left[\log \left(\pi e \left(\sigma_n^2 + \sigma_{e_{\mathrm{int}}}^2 (\mathbf{x}_{\backslash 0}) | x_0 |^2 \right) \right) \right] \right]. \qquad (6.51)$$

Notice, the interpolation error variance $\sigma_{e_{\mathrm{int}}}^2 (\mathbf{x}_{\backslash 0})$ given in (6.37) depends on the distribution of the input symbols $\mathbf{x}_{\backslash 0}$.

6.2.5.2 Upper Bound on $h(y_0 | \hat{h}_0, \mathbf{x}_{\backslash 0})$

Now, we discuss the first term on the RHS of (6.49)

$$h(y_0 | \hat{h}_0, \mathbf{x}_{\backslash 0}) = h(y_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}). \qquad (6.52)$$

In this expression, we do not condition on x_0. Thus, here y_0 is not proper Gaussian. The expectation and the variance are given by

$$\mathrm{E} \left[y_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0} \right] = \mathrm{E} \left[h_0 x_0 + n_0 | \mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0} \right] = 0 \qquad (6.53)$$

$$\begin{aligned}
\text{var}\left[y_0|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] &= \text{E}\left[|y_0|^2|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] \\
&= \text{E}\left[|h_0 x_0 + n_0|^2|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] \\
&= \sigma_x^2 \text{E}\left[|h_0|^2|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] + \sigma_n^2 \\
&= \sigma_x^2 \text{E}\left[|\hat{h}_0 + e_0|^2|\mathbf{x}_{\backslash 0}, \mathbf{y}_{\backslash 0}\right] + \sigma_n^2 \\
&= \sigma_x^2\left(|\hat{h}_0|^2 + \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})\right) + \sigma_n^2.
\end{aligned} \tag{6.54}$$

As proper Gaussian random variables maximize entropy [82], $h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ can be upper-bounded by

$$\begin{aligned}
h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) &\leq \text{E}_{\mathbf{x}_{\backslash 0}}\left[\text{E}_{\hat{h}_0}\left[\log\left(\pi e\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2|\hat{h}_0|^2\right)\right)|\mathbf{x}_{\backslash 0}\right]\right] \\
&= \text{E}_{\mathbf{x}_{\backslash 0}}\left[\int_0^\infty \log\left(\pi e\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2(\sigma_h^2 - \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}))z\right)\right) e^{-z}dz\right].
\end{aligned} \tag{6.55}$$

6.2.5.3 Derivation of an Upper Bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$

Inserting (6.55) and (6.51) into (6.49) yields the following upper bound

$$\begin{aligned}
&\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) \\
&\leq \text{E}_{\mathbf{x}_{\backslash 0}}\left[\int_0^\infty \log\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2(\sigma_h^2 - \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}))z\right) e^{-z}dz\right. \\
&\qquad\qquad \left. - \text{E}_{x_0}\left[\log\left(\sigma_n^2 + \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})|x_0|^2\right)\right]\right].
\end{aligned} \tag{6.56}$$

Obviously, until now we have not discussed the channel interpolation error variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$. Like the channel prediction error variance, the channel interpolation error variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$ is a random variable. Its distribution depends on the distribution of the past and future transmit symbols, which are contained in the vector $\mathbf{x}_{\backslash 0}$. It is difficult to give a general statement on the behavior of (6.56) with respect to $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$. Therefore, in the following we will discuss two cases. On the one hand, the case that the data symbols are constant modulus symbols, and on the other hand, the case that the data symbols are zero-mean proper Gaussian.

Constant Modulus Input Distribution

For a constant modulus (CM) input distribution, i.e., $|x_0|^2 = \sigma_x^2$, the upper bound in (6.56) becomes

$$\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\mathrm{CM}} \leq \int_0^\infty \log\left(1 + \frac{\sigma_x^2(\sigma_h^2 - \sigma_{e_{\mathrm{int,CM}}}^2)}{\sigma_n^2 + \sigma_x^2\sigma_{e_{\mathrm{int,CM}}}^2}z\right)e^{-z}dz \qquad (6.57)$$

where $\sigma_{e_{\mathrm{int,CM}}}^2$ is the channel interpolation error variance in case all past and all future transmit symbols are constant modulus symbols. For the case of an infinite past and future time horizon this estimation variance is no longer random and can be given in closed form, see (6.75) in Section 6.2.5.4. Thus, for this case (6.57) is already an upper bound on the achievable rate with constant modulus symbols in closed form, which can be easily evaluated.

Proper Gaussian Input Distribution

For a zero-mean proper Gaussian distributed x_0 the upper bound in (6.56) becomes

$$\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\mathrm{PG}}$$

$$\leq \mathrm{E}_{\mathbf{x}_{\backslash 0}}\left[\int_0^\infty \log\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2(\sigma_h^2 - \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}))z\right)e^{-z}dz\right.$$

$$\left. - \int_{z=0}^\infty \log\left(\sigma_n^2 + \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})\sigma_x^2 z\right)e^{-z}dz\right]. \qquad (6.58)$$

It can be shown that the RHS of (6.58) monotonically decreases with $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$, see Appendix A.8. However, as the channel interpolation error variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ is itself a random quantity, this statement does not help us to construct an upper bound on the achievable rate. Instead, as in Chapter 4, we would have to construct an upper bound on the RHS of (6.58) by finding the input distribution of the past and future transmit symbols contained in $\mathbf{x}_{\backslash 0}$ among all distributions with i.i.d. input symbols and a maximum average power of σ_x^2, which leads to a distribution of the channel interpolation error variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ that maximizes the RHS of (6.58).

6.2.5.4 The Interpolation Error Variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$

The interpolation error variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ depends on the distribution of all input symbols in the past and in the future, i.e., $\mathbf{x}_{\backslash 0}$. To upper-bound the RHS of (6.58) we, therefore, would need to find the distribution of the input symbols $\mathbf{x}_{\backslash 0}$ among all i.i.d. input distributions with average power σ_x^2, which maximizes the RHS of (6.58).

Similar to Section 4.3.1 for the case of channel prediction, we express the interpolation error variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ as a function of the past and future transmit symbols. For the case of a finite past and future time horizon we

get

$$\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) = \sigma_h^2 - \mathbf{r}_{\mathbf{y}_{\backslash 0}h_0|\mathbf{x}_{\backslash 0}}^H \mathbf{R}_{\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}}^{-1} \mathbf{r}_{\mathbf{y}_{\backslash 0}h_0|\mathbf{x}_{\backslash 0}} \tag{6.59}$$

where $\mathbf{R}_{\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}}$ is the correlation matrix of the observations $\mathbf{y}_{\backslash 0}$ while the past and future transmit symbols $\mathbf{x}_{\backslash 0}$ are known, i.e.,

$$\begin{aligned}
\mathbf{R}_{\mathbf{y}_{\backslash 0}|\mathbf{x}_{\backslash 0}} &= \mathrm{E}\left[\mathbf{y}_{\backslash 0}(\mathbf{y}_{\backslash 0})^H|\mathbf{x}_{\backslash 0}\right] \\
&= \mathrm{E}\left[\left(\mathbf{X}_{\backslash 0}\mathbf{h}_{\backslash 0} + \mathbf{n}_{\backslash 0}\right)\left(\mathbf{X}_{\backslash 0}\mathbf{h}_{\backslash 0} + \mathbf{n}_{\backslash 0}\right)^H|\mathbf{x}_{\backslash 0}\right] \\
&= \mathbf{X}_{\backslash 0}\mathbf{R}_{h\backslash 0}\mathbf{X}_{\backslash 0}^H + \sigma_n^2\mathbf{I}_{2N}
\end{aligned} \tag{6.60}$$

with $\mathbf{X}_{\backslash 0}$ being a diagonal matrix containing the past and future transmit symbols such that

$$\mathbf{X}_{\backslash 0} = \mathrm{diag}\left(\mathbf{x}_{\backslash 0}\right). \tag{6.61}$$

In addition, the vectors $\mathbf{h}_{\backslash 0}$ and $\mathbf{n}_{\backslash 0}$ contain the channel fading weights and the noise samples for all time instances except of the time instant 0. The matrix $\mathbf{R}_{h\backslash 0}$ is given by

$$\mathbf{R}_{h\backslash 0} = \mathrm{E}\left[\mathbf{h}_{\backslash 0}\mathbf{h}_{\backslash 0}^H\right]. \tag{6.62}$$

The cross correlation vector $\mathbf{r}_{\mathbf{y}_{\backslash 0}h_0|\mathbf{x}_{\backslash 0}}$ between the fading weight h_0 and the observation vector $\mathbf{y}_{\backslash 0}$ while knowing the past transmit symbols $\mathbf{x}_{\backslash 0}$ is given by

$$\begin{aligned}
\mathbf{r}_{\mathbf{y}_{\backslash 0}h_0|\mathbf{x}_{\backslash 0}} &= \mathrm{E}\left[\mathbf{y}_{\backslash 0}h_0^*|\mathbf{x}_{\backslash 0}\right] \\
&= \mathrm{E}\left[\left(\mathbf{X}_{\backslash 0}\mathbf{h}_{\backslash 0} + \mathbf{n}_{\backslash 0}\right)h_0^*|\mathbf{x}_{\backslash 0}\right] \\
&= \mathbf{X}_{\backslash 0}\mathbf{r}_{h,\text{int}}
\end{aligned} \tag{6.63}$$

with

$$\mathbf{r}_{h,\text{int}} = [r_h(-N)\ldots r_h(-1)\ \ r_h(1)\ldots r_h(N)]^T \tag{6.64}$$

where $r_h(l)$ is the autocorrelation function as defined in (2.2).

Introducing (6.60) and (6.63) into (6.59) results in

$$\begin{aligned}
\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) &= \sigma_h^2 - \mathbf{r}_{h,\text{int}}^H\mathbf{X}_{\backslash 0}^H\left(\mathbf{X}_{\backslash 0}\mathbf{R}_{h\backslash 0}\mathbf{X}_{\backslash 0}^H + \sigma_n^2\mathbf{I}_{2N}\right)^{-1}\mathbf{X}_{\backslash 0}\mathbf{r}_{h,\text{int}} \\
&= \sigma_h^2 - \mathbf{r}_{h,\text{int}}^H\left(\mathbf{R}_{h\backslash 0} + \sigma_n^2\left(\mathbf{X}_{\backslash 0}^H\mathbf{X}_{\backslash 0}\right)^{-1}\right)^{-1}\mathbf{r}_{h,\text{int}} \\
&\stackrel{(a)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{int}}^H\left(\mathbf{R}_{h\backslash 0} + \sigma_n^2\mathbf{Z}_{\backslash 0}^{-1}\right)^{-1}\mathbf{r}_{h,\text{int}}
\end{aligned} \tag{6.65}$$

where with (a) $\mathbf{Z}_{\backslash 0}$ is defined by

$$\mathbf{Z}_{\backslash 0} = \mathbf{X}_{\backslash 0}^H \mathbf{X}_{\backslash 0} \tag{6.66}$$

i.e., $\mathbf{Z}_{\backslash 0}$ is a diagonal matrix containing the powers of the individual transmit symbols in the past and in the future. Obviously, the channel interpolation error variance is independent of the phases of the past and the future transmit symbols, but only depends on their individual powers. Therefore, we can rewrite (6.58) as

$$\mathcal{I}(x_0; y_0 | \hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\mathrm{PG}}$$
$$\leq \mathrm{E}_{\mathbf{z}_{\backslash 0}} \Bigg[\int_0^\infty \log \left(\sigma_n^2 + \sigma_x^2 \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0}) + \sigma_x^2 (\sigma_h^2 - \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0})) z \right) e^{-z} dz$$
$$- \int_{z=0}^\infty \log \left(\sigma_n^2 + \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0}) \sigma_x^2 z \right) e^{-z} dz \Bigg] \tag{6.67}$$

with $\sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0}) = \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ and where the vector $\mathbf{z}_{\backslash 0}$ corresponds to the diagonal of the matrix $\mathbf{Z}_{\backslash 0}$.

For the derivation of an upper bound on the RHS of (6.67) we would have to maximize the RHS of (6.67) over all distributions of the individual elements of $\mathbf{z}_{\backslash 0}$, i.e., the powers of the individual transmit symbols, with a given maximum average power σ_x^2. Here it has to be considered that we assume that the elements of $\mathbf{z}_{\backslash 0}$ are i.i.d.. If it would be possible to show that the argument of the expectation operation on the RHS of (6.67), i.e.,

$$\int_0^\infty \log \left(\sigma_n^2 + \sigma_x^2 \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0}) + \sigma_x^2 (\sigma_h^2 - \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0})) z \right) e^{-z} dz$$
$$- \int_{z=0}^\infty \log \left(\sigma_n^2 + \sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0}) \sigma_x^2 z \right) e^{-z} dz$$
$$= \int_0^\infty \log \left(\sigma_n^2 + \sigma_x^2 \sigma_h^2 z + \sigma_x^2 (1-z) \left(\sigma_h^2 - \mathbf{r}_{h,\mathrm{int}}^H \left(\mathbf{R}_{h\backslash 0} + \sigma_n^2 \mathbf{Z}_{\backslash 0}^{-1} \right)^{-1} \mathbf{r}_{h,\mathrm{int}} \right) \right) e^{-z} dz$$
$$- \int_{z=0}^\infty \log \left(\sigma_n^2 + \sigma_x^2 z \left(\sigma_h^2 - \mathbf{r}_{h,\mathrm{int}}^H \left(\mathbf{R}_{h\backslash 0} + \sigma_n^2 \mathbf{Z}_{\backslash 0}^{-1} \right)^{-1} \mathbf{r}_{h,\mathrm{int}} \right) \right) e^{-z} dz \tag{6.68}$$

is concave with respect to each individual element of the diagonal of $\mathbf{Z}_{\backslash 0}$ and using Jensen's inequality, this would mean that the RHS of (6.67) is maximized in case the channel interpolation variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0})$ is calculated under the assumption that all past and future transmit symbols are constant modulus symbols with power σ_x^2. This approach would be analogous to the channel prediction case discussed in Section 4.3.1. Unfortunately, we have not been able to prove the concavity of the RHS of (6.68) with respect to the individual elements of $\mathbf{z}_{\backslash 0}$. In contrast to the upper bound based on

channel prediction discussed in Section 4.3.1, evaluation of the concavity of the RHS of (6.68) is more difficult, as the channel interpolation variance is also contained in the first integral of (6.68), corresponding to the upper bound on $h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$.

Nevertheless, for small channel dynamics it is reasonable to approximate the channel interpolation error variance $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ by the channel estimation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols. Therefore, consider that in case of small channel dynamics, the calculation of the channel estimate corresponds to a weighted averaging of many channel output observations. Furthermore, it can easily be shown that in case of a constant channel the distribution of the past and future input symbols is irrelevant, only their average power has an influence on $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$. Thus, we are able to approximate $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ by assuming that all past and future transmit symbols are constant modulus symbols with power σ^2_x. The advantage of this specific assumption on the distribution of the past and future transmit symbols is that, when additionally considering an infinite long observation horizon in the past and in the future, we are able to give a closed form expression for $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$.

For an infinitely long time horizon in the past and in the future, as it has to be considered when calculating the mutual information rate $\mathcal{I}'(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\text{PG}}$ corresponding to $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\text{PG}}$, see (6.67), the distribution of the channel interpolation error variance $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ at each individual time instant is equal. In addition, for the case that all past and all future transmit symbols are assumed to be constant modulus symbols, the channel interpolation error is a constant for all time instances. Therefore, for the channel interpolation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols, i.e., the elements of $\mathbf{x}_{\backslash 0}$ are constant modulus symbols with power $|x_k|^2 = \sigma^2_x$, and an infinitely long past and future time horizon we introduce the following notation

$$\sigma^2_{e_{\text{int}},\text{CM},\infty} = \lim_{N \to \infty} \mathrm{E}\left[\left| h_0 - \hat{h}_0 \right|^2 \bigg| \mathbf{x}_{\backslash 0} : \text{CM}, \mathbf{y}_{\backslash 0} \right]. \qquad (6.69)$$

The index CM denotes constant modulus.

The assumption on constant modulus transmit symbols is obviously in contrast to the assumption on i.i.d. zero-mean proper Gaussian input symbols. However, we use the assumption on constant modulus input symbols only in the context of the channel interpolation error variance $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$.

In addition, the assumption on i.i.d. constant modulus input symbols to calculate $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ is not a severe restriction due to the following reasons. First, the case of using pilot symbols is included, as the pilot symbols are also assumed to be constant modulus symbols with a power of σ^2_x. Secondly, this unveils also that the approximate upper bound calculated under the

assumption that $\sigma_{e_{\mathrm{int}}}^2(\mathbf{z}_{\backslash 0})$ is calculated for past and future data symbols with constant modulus also holds for the special case of using pilot symbols.

Now we calculate the channel interpolation error variance $\sigma_{e_{\mathrm{int},\mathrm{CM},\infty}}^2$. As the problem is jointly Gaussian, the MMSE estimator is linear and the filter input is given by

$$\tilde{y}_k = h_k + \frac{n_k}{x_k}. \tag{6.70}$$

As we assume constant modulus input symbols and as the noise samples n_k are i.i.d. and circularly symmetric, we can rewrite (6.70) as

$$\tilde{y}_k = h_k + \frac{\tilde{n}_k}{\sigma_x} \tag{6.71}$$

where $\tilde{n}_k$ has the same statistical properties as n_k. The PSD of the process $\{\tilde{y}_k\}$ is given by

$$S_{\tilde{y}}(f) = S_h(f) + \frac{\sigma_n^2}{\sigma_x^2}. \tag{6.72}$$

As the channel interpolation error for a process which is characterized by the PSD $S_{\tilde{y}}(f)$ is in general given by [90]

$$\sigma_{e_{\mathrm{int},\tilde{y},\infty}}^2 = \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} [S_{\tilde{y}}(f)]^{-1}\, df \right\}^{-1} \tag{6.73}$$

we get for the interpolation error variance of $\tilde{y}_0$

$$\sigma_{e_{\mathrm{int},\tilde{y},\infty}}^2 = \frac{\sigma_n^2}{\sigma_x^2} \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1 \right]^{-1} df \right\}^{-1}. \tag{6.74}$$

The interpolation error variance of h_0 is then given by subtraction of σ_n^2/σ_x^2 from $\sigma_{e_{\mathrm{int},\tilde{y},\infty}}^2$, yielding

$$\sigma_{e_{\mathrm{int},\mathrm{CM},\infty}}^2 = \frac{\sigma_n^2}{\sigma_x^2} \left[\left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1 \right]^{-1} df \right\}^{-1} - 1 \right]. \tag{6.75}$$

For the special case of a rectangular PSD, (6.75) can be explicitly given as

$$\sigma_{e_{\mathrm{int},\mathrm{CM},\infty}}^2 = \frac{\sigma_h^2}{\rho\left(\frac{1}{2f_d} - 1\right) + 1} \tag{6.76}$$

which is obviously zero for $f_d \to 0$. On the other hand, for $f_d = \frac{1}{2}$ (6.76) becomes

$$\sigma^2_{e_{\text{int}},\text{CM},\infty}\big|_{f_d=\frac{1}{2}} = \sigma^2_h. \tag{6.77}$$

Notice, it should be quite obvious that the interpolation error variance $\sigma^2_{e_{\text{int}},\text{CM},\infty}$ is in general smaller than the prediction error variance $\sigma^2_{e_{\text{pred}},\text{CM},\infty}$ given in (4.53). For a proof of this, we refer to Appendix A.10. This also causes that the *channel interpolation separation* of mutual information introduced in Section 6.2.2 leads to an upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$.

6.2.6 Approximative Upper Bound on the Achievable Rate with the Iterative Code-Aided Synchronized Detection based Receiver using the Modified Channel Estimator (6.19)

With (6.58) and (6.75) we get the following approximate upper bound

$$\mathcal{I}'(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\text{PG}}$$
$$\lesssim \int_0^\infty \log\left(\sigma^2_n + \sigma^2_x\sigma^2_{e_{\text{int}},\text{CM},\infty} + \sigma^2_x(\sigma^2_h - \sigma^2_{e_{\text{int}},\text{CM},\infty})z\right)e^{-z}dz$$
$$- \int_{z=0}^\infty \log\left(\sigma^2_n + \sigma^2_{e_{\text{int}},\text{CM},\infty}\sigma^2_x z\right)e^{-z}dz$$
$$= \int_0^\infty \log\left(1 + \rho\frac{\sigma^2_{e_{\text{int}},\text{CM},\infty}}{\sigma^2_h} + \rho\left(1 - \frac{\sigma^2_{e_{\text{int}},\text{CM},\infty}}{\sigma^2_h}\right)z\right)e^{-z}dz$$
$$- \int_{z=0}^\infty \log\left(1 + \rho\frac{\sigma^2_{e_{\text{int}},\text{CM},\infty}}{\sigma^2_h}z\right)e^{-z}dz$$
$$= \mathcal{I}'_U(y_0; x_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\text{PG}}. \tag{6.78}$$

Based on the discussions in Section 6.2.1 - Section 6.2.5, this is an approximate upper bound on the achievable rate with the receiver described by (6.19), (6.15), and (6.17), i.e, using iterative code-aided synchronized detection with the modified channel estimator presented in Section 6.1.1 and i.i.d. zero-mean proper Gaussian data-symbols.

Notice, if we could prove that the calculation of $\sigma^2_{e_{\text{int}}}(\mathbf{x}_{\backslash 0})$ under the assumption that $\mathbf{x}_{\backslash 0}$ are constant modulus symbols maximizes the RHS of (6.58), (6.78) would be a true upper bound on $\mathcal{I}'(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\text{PG}}$.

Notice that this bound holds without any assumption on the use of pilot symbols. In case we use pilot symbols, the achievable rate decreases as the pilot symbols are deterministic. Therefore, for a pilot spacing L the bound in (6.78) can be modified as follows

$$\mathcal{I}'(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\mathrm{PG}}$$

$$\lesssim \frac{L-1}{L}\left\{ \int_0^\infty \log\left(1 + \rho\frac{\sigma_{e_{\mathrm{int,CM},\infty}}^2}{\sigma_h^2} + \rho\left(1 - \frac{\sigma_{e_{\mathrm{int,CM},\infty}}^2}{\sigma_h^2}\right)z\right)e^{-z}dz\right.$$

$$\left. - \int_{z=0}^\infty \log\left(1 + \rho\frac{\sigma_{e_{\mathrm{int,CM},\infty}}^2}{\sigma_h^2}z\right)e^{-z}dz\right\}$$

$$= \mathcal{I}'_{U_{\mathrm{Pil}}}(y_0; x_0|\hat{h}_0, \mathbf{x}_{\backslash 0})\big|_{\mathrm{PG}}. \tag{6.79}$$

We can give no statement on the amount of required pilot symbols, i.e., the pilot spacing L that assures convergence of the iterative code-aided synchronized detection based receiver in Fig. 6.1 to error-free decoding. In contrast to (6.79), the bound in (6.78) holds irrespectively of any assumptions on the pilot spacing L and obviously upper-bounds the bound in (6.79).

6.2.7 Numerical Evaluation

In the present section, we numerically evaluate the new approximate upper bound on the achievable rate with the receiver using iterative synchronization and decoding with the modified channel estimator presented in Section 6.1.1, which is described by (6.19), (6.15), and (6.17), and i.i.d. zero-mean proper Gaussian data input symbols. This approximate upper bound is given in (6.78). Furthermore, we compare it to the general bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs, i.e., the upper bound in (3.84)/(3.87) and the lower bound in (3.93)/(3.94). In addition, we also compare the approximate upper bound in (6.78) with the achievable rate in case of synchronized detection in combination with a solely pilot based channel estimation discussed in Chapter 5, i.e., the lower bound in (5.39) and the upper bound in (5.42). Here, we only consider the case without pilot-to-average data power ratio optimization.

Fig. 6.2 shows the comparison of the different bounds. Obviously, for the important range of small channel dynamics the approximate upper bound on the achievable rate with the iterative synchronization and decoding based receiver given in (6.78) is smaller than the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., without constraining to the iterative code-aided synchronized detection based receiver described by (6.19), (6.15), and (6.17) in Section 6.1. Thus, the approximate upper bound in (6.78) is valuable, as it gives a tighter upper bound for a realistic iterative code-aided synchronized detection based receiver than the one given in (3.84)/(3.87).

As on the other hand both upper bounds on the achievable rate, the general one in (3.84)/(3.87), as well as the approximate one in (6.78) for the iterative synchronization and decoding based receiver, are not tight, we are not able
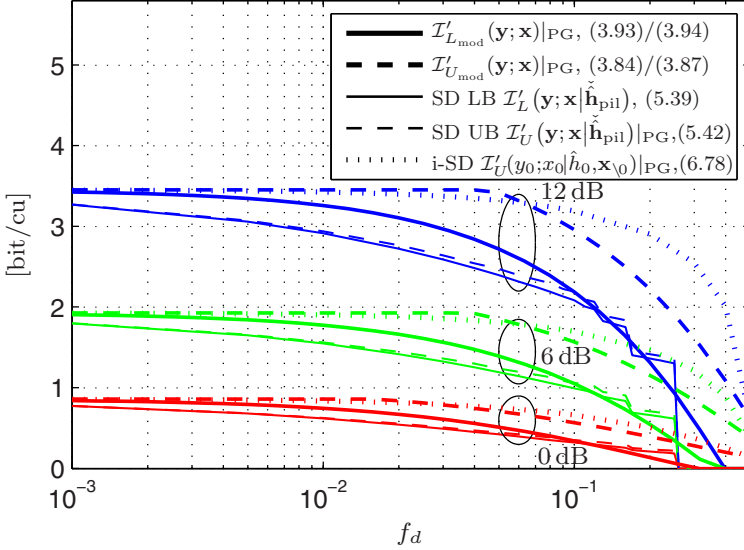
Fig. 6.2: **Approximate upper bound on the achievable rate with the iterative code-aided synchronized detection (i-SD) based receiver using the modified channel estimator (6.19) given in (6.78)** in comparison to bounds on the achievable rate with i.i.d. input symbols and in comparison to bounds on the achievable rate with solely pilot based synchronized detection (SD) without pilot power optimization; in all cases i.i.d. zero-mean proper Gaussian (data) symbols are assumed; rectangular PSD $S_h(f)$

to judge on the actual performance loss due to the restriction to the iterative code-aided synchronized detection based receiver as described in Section 6.1 and based on the channel estimator in (6.19), i.e., the loss being reflected by the fact that the information corresponding to $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0})$ cannot be exploited.

Furthermore, as the iterative synchronization and decoding based approximate upper bound in (6.78) is not tight, it is not directly possible to evaluate the possible gain while using iterative code-aided synchronized detection in comparison to synchronized detection with a solely pilot based channel estimation.

On the other hand, the approximate bound (6.79) enables to evaluate the maximum possible gain that can be achieved by iteratively enhancing the channel estimation using the estimator in (6.19) instead of restricting to a solely pilot based channel estimation in case the pilot spacing is already given. In the following, we choose the maximum pilot spacing fulfilling Nyquist sampling of the channel fading process, i.e., $L = \lfloor 1/(2f_d) \rfloor$. This

pilot spacing is reasonable as, due to Nyquist sampling of the channel fading process, it leads to channel estimates of sufficient quality and, thus, gives good conditions for low bit error rates (BER) after decoding. However, the BER will furthermore depend on the SNR and the specific code and its code rate. In Fig. 6.3 we show the modified approximate upper bound given in (6.79) with $L = \lfloor 1/(2f_d) \rfloor$ in comparison to the bounds on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation. For the latter ones the pilot spacing which maximizes the lower bound on the achievable rate with solely pilot based synchronized detection has been chosen. Notice that this pilot spacing might be smaller than the one corresponding to Nyquist sampling of the channel fading process, see Section 5.2. Furthermore, it should be noticed that the approximate upper bound in (6.79) is not a strict upper bound on the achievable rate, as it might be possible to choose a larger pilot spacing than $L = \lfloor 1/(2f_d) \rfloor$. The maximum possible pilot spacing still assuring convergence to error-free decoding is an open question.

Obviously, for small channel dynamics the approximate upper bound on the achievable rate with the receiver using iterative code-aided synchronized detection is significantly above the bounds on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation. However, consider that we only show an approximate upper bound, therefore, the obvious gap is only an upper bound on the possible gain while using the discussed iterative synchronization and decoding scheme.

To further evaluate the possible performance gain while using iterative synchronization and decoding in comparison to solely pilot based synchronized detection, we would have to find and upper bound that holds also when using the genuine channel estimator in (6.13), which also uses the observation y_k for the estimation of h_k. Additionally, a tight lower bound on the achievable rate with such a receiver structure is required. Clearly, the lower bound on the achievable rate with solely pilot based synchronized detection given in (5.39) is a lower bound on the achievable rate with iterative synchronization and decoding. But obviously it is not tight, as the possible gain due to an enhanced channel estimation quality based on the reliability information on the data symbols is not used. Within this work we do not show a tighter lower bound.

Small Channel Dynamics

If we evaluate the approximate upper bound on the achievable rate with the iterative synchronization and decoding based receiver in (6.78) for $\sigma^2_{e_{\text{int,CM},\infty}} \to 0$, i.e., for asymptotically small channel dynamics, we get

$$\mathcal{I}'_U(y_0; x_0 | \hat{h}_0, \mathbf{x}_{\setminus 0}) \big|_{\text{PG}, \sigma^2_{e_{\text{int,CM},\infty}} = 0} = \int_0^\infty \log\left(1 + \rho z\right) e^{-z} dz \qquad (6.80)$$
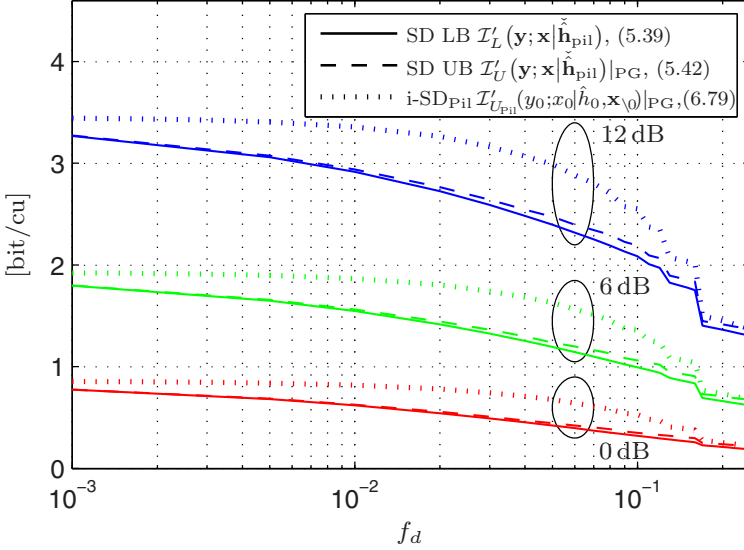
Fig. 6.3: **Approximate upper bound on the achievable rate with the iterative code-aided synchronized detection based receiver using the modified channel estimator (6.19) and a pilot spacing corresponding to Nyquist channel sampling (i-SD$_{\text{Pil}}$) (6.79)** ($L = \lfloor 1/(2f_d) \rfloor$); for comparison: bounds on the achievable rate with solely pilot based synchronized detection (SD) without pilot power optimization; in all cases i.i.d. zero-mean proper Gaussian data symbols are assumed; for solely pilot based SD the pilot spacing that maximizes the lower bound on the achievable rate has been chosen; notice, due to the given pilot spacing L (6.79) is not a strict upper bound on the achievable rate with the iterative code-aided synchronized detection based receiver; rectangular PSD $S_h(f)$

which obviously corresponds to the coherent capacity. This behavior can also be observed in Fig. 6.2.

6.3 Summary

In the present chapter, we have derived an upper bound on the achievable rate with the iterative code-aided synchronized detection based receiver described by (6.19), (6.15), and (6.17) in Section 6.1. Therefore, we have recalled the principle of iterative code-aided synchronized detection. Furthermore, we identified the nature of the possible gain by iteratively enhancing the channel estimation based on reliability information on the data symbols. This pos-

sible gain arises from the utilization of parts of the information contained in the temporal correlation of the channel estimation error process, which have been discarded while using a receiver based on synchronized detection in combination with a solely pilot based channel estimation as described in Chapter 5.

For the specific structure of an iterative code-aided synchronized detection based receiver as described by (6.19), (6.15), and (6.17) we have derived an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian data symbols. Note that this receiver is slightly modified with respect to the typically studied receiver based on iterative code-aided channel estimation, as the channel estimator does not use the observation y_k for the estimation of h_k. The typically studied receiver structure, which is also described in Section 6.1, results from a set of fixed point equations, which is derived based on the joint ML detection and MAP parameter estimation problem. We have shown that this modified receiver cannot exploit the complete mutual information of the transmitter and the receiver, the information corresponding to $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ cannot be retrieved. This is a result of the fact of the combination of a symbolwise detection metric with the channel estimator which does not use y_k for the estimation of h_k. However, for channel dynamics as they are typically observed in mobile environments, we guess that the amount of discarded information due to the modification of the channel estimator is small. We do not know if and which part of the information $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ can be retrieved when using the genuine channel estimator given in (6.13).

Unfortunately, the given upper bound on the achievable rate with the receiver based on iterative code-aided synchronized detection described by (6.19), (6.15), and (6.17) is a non-closed form expression, as it still depends on the channel interpolation error variance for an arbitrarily chosen time instant k, and, thus, is a function of the distribution of all transmit symbols in the past and in the future with respect to the time instant k. We are not able to give a closed form solution. Nevertheless, for small channel dynamics, it is reasonable to approximate the channel interpolation error variance by the channel interpolation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols with power σ_x^2. Based on this approximation we have also numerically evaluated the upper bound. If the statement that the calculation of the channel interpolation error variance under the assumption of constant modulus transmit symbols in the past and in the future yields an upper bound on the RHS of (6.58) could be proven, e.g., by showing the concavity of (6.68) with respect to each individual element of the diagonal of $\mathbf{Z}_{\backslash 0}$, the upper bound in (6.78) would be a true upper bound. In contrast, for the case of constant modulus signaling, we can give a closed form solution for the upper bound on the achievable rate with the discussed receiver based on code-aided iterative synchronized detection.

The approximate upper bound on the achievable rate with the studied receiver based on iterative code-aided synchronized detection described by

(6.19), (6.15), and (6.17) is independent of the used amount of pilot symbols. However, we also have given a modified version of the approximate upper bound taking the rate loss due to pilot symbols into account. This bound can be used to give an approximative upper bound on the maximum possible performance gain for a fixed pilot spacing when using iterative synchronization and decoding instead of a solely pilot based synchronized detection approach as discussed in Chapter 5. But it has to be clearly stated that due to the restrictive assumption on a given pilot spacing this bound is not a true upper bound on the achievable rate with a receiver based on iterative synchronization and decoding, as we do not know how many pilots are required to assure convergence of the iterative receiver to error-free decoding. Furthermore, it holds only when using the modified channel estimator, which does not use y_k for the estimation of h_k. However, we assume that the additional information by using the genuine channel estimator in (6.13) is small in case of typical, i.e., small, channel dynamics.

To further evaluate the performance gain of iterative synchronization and decoding in comparison to synchronized detection used in combination with a solely pilot based channel estimation, it would be necessary to derive an upper bound which holds also when using the genuine channel estimator in (6.13), which results from the set of fixed point equations as described in Section 6.1. Furthermore, the derivation of a tight lower bound on the achievable rate with iterative code-aided channel estimation and synchronized detection is required. However, within this work we do not give such a lower bound. For comparison we only have used the lower bound on the achievable rate with solely pilot based synchronized detection, which is not tight as it does not show the gain due to an enhanced channel estimation based on the reliability information on the data symbols. Recently, for the special case of a block fading channel a lower bound on the achievable rate when using a joint processing of pilot and data symbols has been given in [56]. In the following chapter, we extend this approach to the scenario of a stationary Rayleigh flat-fading channel as it is considered in the present work. However, note that this lower bound on the achievable rate with joint processing of pilot and data symbols is not a lower bound on the achievable rate with the receiver based on synchronized detection and an iterative code-aided channel estimation as discussed in the present chapter, i.e., the receiver described by (6.19), (6.15), and (6.17). Therefor, consider that this form of receiver processing is obviously not optimal, which can be seen from the fact that the information corresponding to the term $\mathcal{I}(y_0; \mathbf{x}_{\setminus 0}|\mathbf{y}_{\setminus 0})$ cannot be exploited by the given receiver structure.

Chapter 7
Joint Processing of Pilot and Data Symbols

In the previous chapter we have discussed the achievable rate with iterative code-aided channel estimation. As it has already been stated there, also this approach is not optimal, as it is based on a coherent detection metric, i.e., the detection is performed symbol-wise, although the channel estimation error is temporally correlated. Therefore, in the present chapter we study the achievable rate with an optimal joint processing of pilot and data symbols. This means that we assume an optimal receiver processing without any restriction to a coherent, i.e., symbol-wise detection. For this *joint processing* of pilot and data symbols we derive a lower bound on the achievable rate. As in the previous chapters we consider a discrete-time Rayleigh flat-fading channel as it is described in Chapter 2. Parts of the present chapter have already been published in [26].

For the case of a joint processing of pilot and data symbols there is not much knowledge on the achievable rate. Very recently, in [56] the value of joint processing of pilot and data symbols has been studied in the context of a block fading channel. To the best of our knowledge, there are no results concerning the gain of joint processing of pilot and data symbols for the case of stationary fading channels. The lower bound on the achievable rate with a joint processing of pilot and data symbols for the case of a stationary Rayleigh flat-fading channel that we give in the present chapter can be seen as an extension of the work for the block fading channel in [56].

Before we derive the lower bound on the achievable rate, we rewrite the mutual information between the transmitter and the receiver based on a channel estimate which is solely based on pilot symbols, as it has already been used in (5.29) in Section 5.2. However, in the present chapter we give a formal derivation of this representation. We give this derivation within the present chapter, as this is an essential part in the derivation of the lower bound on the achievable rate with the joint processing of pilot and data symbols. Finally, we compare the given lower bound on the achievable rate with joint processing of pilot and data symbols to bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in Chapter 3, i.e.,

without the assumption on pilot symbols inserted into the transmit sequence. Furthermore, we compare the lower bound with joint processing of pilot and data symbols to bounds on the achievable rate with synchronized detection based on a solely pilot based channel estimation as it has been discussed in Chapter 5 and with the upper bound on the achievable rate with the specific receiver based on synchronized detection and iterative code-aided channel estimation as it has been discussed in Chapter 6.

7.1 System Model

Concerning the channel, we consider a stationary Rayleigh flat-fading channel with a compactly supported PSD of the channel fading process and $f_d < \frac{1}{2}$ as it is introduced in Chapter 2.

As in Chapter 5.2, for the following derivation we use the subvectors $\mathbf{x}_D$ containing all data symbols of $\mathbf{x}$ and the vector $\mathbf{x}_P$ containing all pilot symbols of $\mathbf{x}$. Correspondingly, we define the vectors $\mathbf{h}_D$, $\mathbf{h}_P$, $\mathbf{y}_D$, $\mathbf{y}_P$, $\mathbf{n}_D$, and $\mathbf{n}_P$.

The transmit symbol sequence consists of data symbols with a maximum average power σ_x^2, i.e.,

$$\frac{1}{N_D}\mathrm{E}\left[\mathbf{x}_D^H\mathbf{x}_D\right] \leq \sigma_x^2 \tag{7.1}$$

with N_D being the length of the vector $\mathbf{x}_D$, and periodically inserted pilot symbols with a fixed transmit power σ_x^2. Each L-th symbol is a pilot symbol. We assume that the pilot spacing is chosen such that the channel fading process is sampled at least with Nyquist rate, i.e.,

$$L < \frac{1}{2f_d}. \tag{7.2}$$

7.2 Expressing $\mathcal{I}(\mathbf{y};\mathbf{x})$ via the Pilot based Channel Estimate $\hat{\mathbf{h}}_{\mathbf{pil}}$

Before we quantitatively discuss the value of a joint processing of data and pilot symbols, we rewrite the mutual information between the transmitter and the receiver based on a solely pilot based channel estimate, as it has already been used in (5.29) in Section 5.2. However, here we show a formal derivation, to which we refer in the following derivation of the lower bound on the achievable rate with a joint processing of pilot and data symbols.

Recall that the mutual information between the transmitter and the receiver is given by $\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)$, see (5.29) (c). As the pilot symbols are

known to the receiver, the pilot symbol vector $\mathbf{x}_P$ is found at the RHS of the semicolon. We separate $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$ as follows

$$\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \mathbf{y}_P, \mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P | \mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{x}_P)$$

$$\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \mathbf{y}_P, \mathbf{x}_P) \tag{7.3}$$

where (a) follows from the chain rule for mutual information and (b) holds due to the independency of the data and pilot symbols. Up to this point the derivation is analogous to Section 5.2, cf. (5.29) (f).

Now, we formally show that the conditioning on $\mathbf{y}_P, \mathbf{x}_P$ in (7.3) is equal to conditioning on a solely pilot based channel estimate $\hat{\mathbf{h}}_{\text{pil}}$, see (5.12), and $\mathbf{x}_P$.

To prove that the condition on $\mathbf{y}_P$ and $\mathbf{x}_P$ in (7.3) can be substituted by $\hat{\mathbf{h}}_{\text{pil}}$ and $\mathbf{x}_P$ let us consider ML sequence detection. The receiver has to find the most likely data sequence $\mathbf{x}_D$ based on the observation $\mathbf{y}$ while knowing the pilot symbols $\mathbf{x}_P$, i.e.,

$$\hat{\mathbf{x}}_D = \arg \max_{\mathbf{x}_D \in \mathcal{P}_{\mathbf{x}_D}} p(\mathbf{y}|\mathbf{x})$$

$$= \arg \max_{\mathbf{x}_D \in \mathcal{P}_{\mathbf{x}_D}} p(\mathbf{y}_D | \mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P) \tag{7.4}$$

with the set $\mathcal{P}_{\mathbf{x}_D}$ containing all data sequences $\mathbf{x}_D$ that can be generated by the transmitter. It can be shown that the PDF $p(\mathbf{y}_D | \mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P)$ is proper Gaussian and, thus, is completely described by the conditional mean and covariance

$$\mathrm{E}\left[\mathbf{y}_D | \mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P\right] = \mathbf{X}_D \mathrm{E}\left[\mathbf{h}_D | \mathbf{y}_P, \mathbf{x}_P\right] = \mathbf{X}_D \hat{\mathbf{h}}_{\text{pil},D} \tag{7.5}$$

$$\mathrm{cov}[\mathbf{y}_D | \mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P] = \mathbf{X}_D \mathbf{R}_{e_{\text{pil}},D} \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_{N_D} \tag{7.6}$$

where $\mathbf{X}_D = \mathrm{diag}(\mathbf{x}_D)$ and $\mathbf{I}_{N_D}$ is an identity matrix of size $N_D \times N_D$. The vector $\hat{\mathbf{h}}_{\text{pil},D}$ is an MMSE channel estimate at the data symbol time instances based on the pilot symbols, which is denoted by the index pil. Due to the jointly proper Gaussian nature of this problem, the MMSE estimate is linear, i.e., it is an LMMSE estimate. Furthermore, the corresponding channel estimation error

$$\mathbf{e}_{\text{pil},D} = \mathbf{h}_D - \hat{\mathbf{h}}_{\text{pil},D} \tag{7.7}$$

is zero-mean proper Gaussian and

$$\mathbf{R}_{e_{\text{pil}},D} = \mathrm{E}\left[\mathbf{e}_{\text{pil},D} \mathbf{e}_{\text{pil},D}^H | \mathbf{x}_P\right] \tag{7.8}$$

is its correlation matrix, which is independent of $\mathbf{y}_P$ due to the principle of orthogonality in LMMSE estimation.

Based on (7.5) and (7.6) conditioning of $\mathbf{y}_D$ on $\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P$ is equivalent to conditioning on $\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P$, i.e.,

$$p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P) = p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P) \qquad (7.9)$$

as all information on $\mathbf{h}_D$ delivered by $\mathbf{y}_P$ is contained in $\hat{\mathbf{h}}_{\mathrm{pil},D}$ while conditioning on $\mathbf{x}_P$. See (4.12) in Section 4.1 and (6.41) in Section 6.2.3 for analogous derivations in the context of channel prediction and interpolation.

Thus, (7.4) can be written as

$$\hat{\mathbf{x}}_D = \arg \max_{\mathbf{x}_D \in \mathcal{P}_{\mathbf{x}_D}} p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P)$$

$$= \arg \max_{\mathbf{x}_D \in \mathcal{P}_{\mathbf{x}_D}} p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P). \qquad (7.10)$$

For ease of notation in the following we will use the metric on the RHS of (7.10) where $\hat{\mathbf{h}}_{\mathrm{pil}}$ corresponds to $\hat{\mathbf{h}}_{\mathrm{pil},D}$ but also contains channel estimates at the pilot symbol time instances, i.e.,

$$\hat{\mathbf{h}}_{\mathrm{pil}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P\right]. \qquad (7.11)$$

Note that the solely pilot based channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$ in (7.11) exactly corresponds to the solely pilot based channel estimate in Chapter 5, see (5.12).

Based on $\hat{\mathbf{h}}_{\mathrm{pil}}$, the genuine input/output relation of the channel given in (2.13) can be expressed by, cf. (5.10)

$$\mathbf{y} = \mathbf{X}(\hat{\mathbf{h}}_{\mathrm{pil}} + \mathbf{e}_{\mathrm{pil}}) + \mathbf{n} \qquad (7.12)$$

where $\mathbf{e}_{\mathrm{pil}}$ is the estimation error including the pilot symbol time instances, i.e.

$$\mathbf{e}_{\mathrm{pil}} = \mathbf{h} - \hat{\mathbf{h}}_{\mathrm{pil}}. \qquad (7.13)$$

As the channel estimation is an interpolation, the error process is not white but temporally correlated, i.e.,

$$\mathbf{R}_{e_{\mathrm{pil}}} = \mathrm{E}\left[\mathbf{e}_{\mathrm{pil}}\mathbf{e}_{\mathrm{pil}}^H|\mathbf{x}_P\right] \qquad (7.14)$$

is not diagonal, cf. (7.25). As the estimation error process is zero-mean proper Gaussian the PDF in (7.10) is given by, cf. (5.24)

$$p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) = \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{X}\mathbf{R}_{e_{\mathrm{pil}}}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right) \qquad (7.15)$$

where $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes a proper Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ and where $\mathbf{I}_N$ is the $N \times N$ identity matrix.[1]

Corresponding to (7.9), we can also rewrite $p(\mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P)$ as follows

$$
\begin{aligned}
p(\mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) &= \int p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P)p(\mathbf{x}_D|\mathbf{y}_P, \mathbf{x}_P)d\mathbf{x}_D \\
&\stackrel{(a)}{=} \int p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P)p(\mathbf{x}_D)d\mathbf{x}_D \\
&= p(\mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P)
\end{aligned}
\tag{7.16}
$$

where for (a) we have used (7.9) and the independency of $\mathbf{x}_D$ of $\mathbf{x}_P$ and $\mathbf{y}_P$. See (6.42) in Section 6.2.3 for an analogous derivation in the context of the channel interpolation.

Based on (7.9) and (7.16), we can also rewrite (7.3) as

$$
\begin{aligned}
\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) &= \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) \\
&\stackrel{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}})
\end{aligned}
\tag{7.17}
$$

where (a) holds as the pilot symbols are deterministic.

Thus we have given a formal derivation of equality (5.29). We show this derivation including the definition of the solely pilot based channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$ in the context of the study of the achievable rate with a joint processing of pilot and data symbols, as it is used in the following derivation of a lower bound on the achievable rate with joint processing of pilot and data symbols.

Obviously, as it has already been stated in Section 5.2, $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}})$ is still equal to the complete mutual information between the transmitter and the receiver. The receiver based on synchronized detection and a solely pilot based channel estimation only uses parts of this information, as the temporal correlation contained in the channel estimation error $\mathbf{e}_{\mathrm{pil}}$ is not exploited by the detector, see the discussion in Section 5.2.

7.3 Lower Bound on the Achievable Rate with Joint Processing

Now, we give a new lower bound on the achievable rate for a joint processing of data and pilot symbols. The following approach can be seen as an extension of the work in [56] for the case of a block fading channel to the stationary Rayleigh flat-fading scenario discussed in the present work. Therefore, analogously to [56] we decompose and lower-bound the mutual information

---

[1] Note that for the case of data transmission only (7.15) becomes $p(\mathbf{y}|\mathbf{x}_D) = \mathcal{CN}(\mathbf{0}, \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2\mathbf{I}_N)$ as in this case $\hat{\mathbf{h}}_{\mathrm{pil}} = \mathbf{0}$ and $\mathbf{R}_{e_{\mathrm{pil}}} = \mathbf{R}_h$, cf. (5.3) in Section 5.1.

between the transmitter and the receiver $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$ as follows

$$
\begin{aligned}
\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) &\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P, \mathbf{h}) - \mathcal{I}(\mathbf{x}_D; \mathbf{h}|\mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \\
&\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h}|\mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \\
&\overset{(c)}{\geq} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)
\end{aligned}
$$
(7.18)

where (a) follows from the chain rule for mutual information. For the first term in (b) we have used the fact that due to the knowledge on $\mathbf{h}$, the knowledge on $\mathbf{y}_P$ and $\mathbf{x}_P$ does not increase the mutual information between $\mathbf{x}_D$ and $\mathbf{y}_D$. Finally, (c) is due to the fact that conditioning reduces entropy. Note, the first term on the RHS of (7.18) is the mutual information in case of perfect channel knowledge.

In the following we deviate from the derivation given in [56]. Now, we calculate both differential entropy terms at the RHS of (7.18). Therefore, we rewrite the RHS of (7.18) as follows

$$
\begin{aligned}
\mathcal{I}(\mathbf{x}_D; \mathbf{y}_P, \mathbf{y}_P, \mathbf{x}_P) &\geq \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \\
&\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h}|\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) + h(\mathbf{h}|\hat{\mathbf{h}}_{\mathrm{joint}}, \mathbf{x}_D, \mathbf{x}_P) \\
&\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\hat{\mathbf{h}}_{\mathrm{pil}} + \mathbf{e}_{\mathrm{pil}}|\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) + h(\hat{\mathbf{h}}_{\mathrm{joint}} + \mathbf{e}_{\mathrm{joint}}|\hat{\mathbf{h}}_{\mathrm{joint}}, \mathbf{x}_D, \mathbf{x}_P) \\
&\overset{(c)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{e}_{\mathrm{pil}}|\mathbf{x}_P) + h(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P) \\
&\overset{(d)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \mathrm{E}_{\mathbf{x}_P}\left[\log \det\left(\pi e \mathbf{R}_{e_{\mathrm{pil}}}\right)\right] + \mathrm{E}_{\mathbf{x}_P, \mathbf{x}_D}\left[\log \det\left(\pi e \mathbf{R}_{e_{\mathrm{joint}}}\right)\right] \\
&\overset{(e)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \log \det\left(\mathbf{R}_{e_{\mathrm{pil}}}\right) + \mathrm{E}_{\mathbf{x}_D}\left[\log \det\left(\mathbf{R}_{e_{\mathrm{joint}}}\right)\right]
\end{aligned}
$$
(7.19)

where for the second term in (a) we have substituted the condition on $\mathbf{y}_P$ by $\hat{\mathbf{h}}_{\mathrm{pil}}$, which is possible as the estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$ contains the same information on $\mathbf{h}$ as $\mathbf{y}_P$ while conditioning on $\mathbf{x}_P$. Corresponding to the solely pilot based channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$, based on $\mathbf{x}_D$, $\mathbf{x}_P$, $\mathbf{y}_D$, and $\mathbf{y}_P$, we can calculate the estimate $\hat{\mathbf{h}}_{\mathrm{joint}}$, which is based on data and pilot symbols. Like $\hat{\mathbf{h}}_{\mathrm{pil}}$ this estimate is a MAP estimate, which, due to the jointly Gaussian nature of the problem, is an MMSE estimate, i.e.,

$$
\hat{\mathbf{h}}_{\mathrm{joint}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P, \mathbf{y}_D, \mathbf{x}_D\right].
$$
(7.20)

Thus, for (a) we have substituted the conditioning on $\mathbf{y}_D$ and $\mathbf{y}_P$ by conditioning on $\hat{\mathbf{h}}_{\mathrm{joint}}$ in the third term, as $\hat{\mathbf{h}}_{\mathrm{joint}}$ contains all information on $\mathbf{h}$ that is contained in $\mathbf{y}_D$ and $\mathbf{y}_P$ while $\mathbf{x}_D$ and $\mathbf{x}_P$ are known. For equality (b) we have used for the second term that $\mathbf{h}$ can be expressed as a sum of its estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$ and the estimation error $\mathbf{e}_{\mathrm{pil}}$, cf. (7.12). Analogously, for the third term we used the separation of $\mathbf{h}$ into the estimate $\hat{\mathbf{h}}_{\mathrm{joint}}$ and the

corresponding estimation error $\mathbf{e}_{\text{joint}}$, i.e.,

$$\mathbf{e}_{\text{joint}} = \mathbf{h} - \hat{\mathbf{h}}_{\text{joint}}. \tag{7.21}$$

Equality (c) is due to the fact that the addition of a constant does not change differential entropy and that the estimation error $\mathbf{e}_{\text{pil}}$ is independent of the estimate $\hat{\mathbf{h}}_{\text{pil}}$, and analogously $\mathbf{e}_{\text{joint}}$, which depends on $\mathbf{x}_P$ and $\mathbf{x}_D$, is independent of $\hat{\mathbf{h}}_{\text{joint}}$ due to the orthogonality principle in LMMSE estimation. Finally, (d) follows from the fact that the estimation error processes are zero-mean jointly proper Gaussian. Here the error correlation matrices are given by (7.14) and by

$$\mathbf{R}_{e_{\text{joint}}} = \mathrm{E}\left[\mathbf{e}_{\text{joint}}\mathbf{e}_{\text{joint}}^H|\mathbf{x}_D, \mathbf{x}_P\right]. \tag{7.22}$$

For (e) we have used that the pilot symbols are deterministic. Therefore, the expectation over $\mathbf{x}_P$ in the second and third term can be removed. However, the channel estimation error $\mathbf{e}_{\text{joint}}$ depends on the distribution of the data symbols $\mathbf{x}_D$.

Concerning the third term on the RHS of (7.19) it can be shown that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$, i.e.,

$$h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \lim_{N \to \infty} \frac{1}{N} h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) \tag{7.23}$$

is minimized for a given maximum average transmit power σ_x^2 if the data symbols are constant modulus (CM) symbols with power σ_x^2, see Appendix A.11. Within this proof the restriction to an absolutely summable autocorrelation function $r_h(l)$, see (2.4), is required.

Thus, based on (7.19) a lower bound for the achievable rate with joint processing of data and pilot symbols is given by

$$\mathcal{I}'(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) = \lim_{N \to \infty} \frac{1}{N} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$$

$$\geq \lim_{N \to \infty} \frac{1}{N} \left\{ \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \log\det\left(\mathbf{R}_{e_{\text{pil}}}\right) + \log\det\left(\mathbf{R}_{e_{\text{joint,CM}}}\right) \right\}$$

$$\overset{(a)}{=} \lim_{N \to \infty} \frac{1}{N} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)}\right) df \tag{7.24}$$

with $\mathbf{R}_{e_{\text{joint,CM}}}$ corresponding to (7.22), but under the assumption of CM data symbols with transmit power σ_x^2. As $\mathbf{R}_{e_{\text{joint,CM}}}$ only depends on the distribution of the magnitude of the data symbols contained in $\mathbf{x}_D$, which is constant and deterministic, we can remove the expectation operation with respect to $\mathbf{x}_D$. Note that the CM assumption has only been used to lower-bound the third term at the RHS of (7.19), and not the whole expression at the RHS of (7.19). For (a) in (7.24) we have used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [44]. $S_{e_{\text{pil}}}(f)$ and

$S_{e\text{joint,CM}}(f)$ are the PSDs of the channel estimation error processes, on the one hand, if the estimation is solely based on pilot symbols, and on the other hand, if the estimation is based on data and pilot symbols, assuming CM data symbols. They are given by

$$S_{e_{\mathrm{pil}}}(f) = \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} \tag{7.25}$$

$$S_{e_{\mathrm{joint,CM}}}(f) = \frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2} + 1}. \tag{7.26}$$

The derivation of these PSDs is given in Appendix A.12.

However, the application of Szegö's theorem for (a) in (7.24) requires several steps, which we discuss in the following. The limit over the second and the third term on the LHS of (a) in (7.24) can be transformed as follows

$$\lim_{N\to\infty} \frac{1}{N} \left\{ \log\det\left(\mathbf{R}_{e_{\mathrm{pil}}}\right) - \log\det\left(\mathbf{R}_{e_{\mathrm{joint,CM}}}\right) \right\}$$

$$\stackrel{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \left\{ \log\det\left(\mathbf{C}_{e_{\mathrm{pil}}}\right) - \log\det\left(\mathbf{C}_{e_{\mathrm{joint,CM}}}\right) \right\}$$

$$\stackrel{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \left\{ \log\det\left(\mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\mathbf{F}^H\right) - \log\det\left(\mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}\mathbf{F}^H\right) \right\}$$

$$= \lim_{N\to\infty} \frac{1}{N} \left\{ \log\det\left(\mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}^{-1}\mathbf{F}^H\right) \right\}$$

$$\stackrel{(c)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_{e_{\mathrm{pil}}}(f)}{S_{e_{\mathrm{joint,CM}}}(f)}\right) df$$

$$\stackrel{(d)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\rho\frac{S_h(f)}{\sigma_h^2} + 1}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1}\right) df \tag{7.27}$$

where for (a) we have substituted the Toeplitz matrices $\mathbf{R}_{e_{\mathrm{pil}}}$ and $\mathbf{R}_{e_{\mathrm{joint,CM}}}$ by their asymptotic equivalent circulant matrices $\mathbf{C}_{e_{\mathrm{pil}}}$ and $\mathbf{C}_{e_{\mathrm{joint,CM}}}$, see [43]. Furthermore, for (b) we have used the spectral decompositions of the circulant matrices given by

$$\mathbf{C}_{e_{\mathrm{pil}}} = \mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\mathbf{F}^H \tag{7.28}$$

$$\mathbf{C}_{e_{\mathrm{joint,CM}}} = \mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}\mathbf{F}^H \tag{7.29}$$

where $\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}$ and $\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}$ are diagonal matrices containing the eigenvalues of $\mathbf{C}_{e_{\mathrm{pil}}}$ and $\mathbf{C}_{e_{\mathrm{joint,CM}}}$, and the matrix $\mathbf{F}$ is a unitary DFT-matrix whose elements are given by

$$[\mathbf{F}]_{k,l} = \frac{1}{\sqrt{N}}e^{j2\pi\frac{(k-1)(l-1)}{N}}. \tag{7.30}$$

For (c) in (7.27) we have then used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [44]. Therefore, first consider that the matrix $\mathbf{F}\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}^{-1}\mathbf{F}^H$ on the LHS of (c) is again a circulant matrix and that there exists an asymptotically equivalent Toeplitz matrix. Furthermore, the eigenvalues of $\mathbf{C}_{e_{\mathrm{pil}}}$ are samples of the PSD $S_{e_{\mathrm{pil}}}(f)$ and the eigenvalues of $\mathbf{C}_{e_{\mathrm{joint,CM}}}$ are samples of the PSD $S_{e_{\mathrm{joint,CM}}}(f)$. Here we assume a construction of the circulant matrices as described in [43, (4.32)], which holds for absolutely summable autocorrelation functions, see (2.4). Furthermore, the application of Szegö's theorem requires that the log-function is continuous on the support of the eigenvalues of the matrix $\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}^{-1}$. This means that we have to show that the eigenvalues of $\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}^{-1}$ are bounded away from zero and from infinity. That this is indeed the case will become obvious after introducing $S_{e_{\mathrm{pil}}}(f)$ and $S_{e_{\mathrm{joint,CM}}}(f)$ given in (7.25) and (7.26) as it has been done in (d). Obviously, the argument of the log at the RHS of (7.27) is larger than zero and smaller than infinity on the interval $f \in [-0.5, 0.5]$. Therefore, the integral on the RHS of (7.27) exists, implying that also the LHS of (c) in (7.27) is bounded and, thus, that the eigenvalues of $\boldsymbol{\Lambda}_{e_{\mathrm{pil}}}\boldsymbol{\Lambda}_{e_{\mathrm{joint,CM}}}^{-1}$ are bounded away from zero and from infinity. Thus, in conclusion we have shown that Szegö's theorem is applicable and that (a) in (7.24) holds.

The first term on the RHS of (7.24) is the mutual information rate in case of perfect channel state information

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) = \lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{h})$$
$$= \lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{h}_D). \tag{7.31}$$

For an average power constraint as given in (7.1), (7.31) is maximized by i.i.d. zero-mean proper Gaussian (PG) data symbols, i.e.,

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{h}_D)\big|_{\mathbf{x}_D\sim\text{i.i.d. PG}} = \frac{L-1}{L}\mathcal{I}(x_{D_k};y_{D_k}|h_{D_k})\big|_{\text{PG}}$$
$$= \frac{L-1}{L}\mathrm{E}_{h_{D_k}}\left[\log\left(1+\rho\frac{|h_{D_k}|^2}{\sigma_h^2}\right)\right]$$
$$= \frac{L-1}{L}\int_{z=0}^{\infty}\log\left(1+\rho z\right)e^{-z}dz$$
$$= \frac{L-1}{L}C_{\mathrm{perf}}(\rho) \tag{7.32}$$

where the subscript D_k denotes an arbitrary data symbol time instant, the factor $(L-1)/L$ arises as each L-th symbol is a pilot symbol, and $C_{\mathrm{perf}}(\rho)$ corresponds to the coherent capacity, cf. (3.89).

Thus, with (7.24), (7.25), (7.26), (7.31) and (7.32) we get the following lower bound on the achievable rate with joint processing

$$\mathcal{R}_{L,\mathrm{joint}} = \frac{L-1}{L} C_{\mathrm{perf}}(\rho) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left(\frac{\rho \frac{S_h(f)}{\sigma_h^2} + 1}{\frac{\rho}{L} \frac{S_h(f)}{\sigma_h^2} + 1} \right) df \qquad (7.33)$$

This lower bound is achievable with i.i.d. zero-mean proper Gaussian data symbols.

7.3.1 Fixed Pilot Spacing

With (7.33) we have found a lower bound on the achievable rate with joint processing of data and pilot symbols, for a given pilot spacing L and stationary Rayleigh flat-fading.

For the special case of a rectangular PSD[2] of the channel fading process as defined in (2.8) the lower bound in (7.33) becomes

$$\mathcal{R}_{L,\mathrm{joint}}\Big|_{\mathrm{rect}.S_h(f)} = \frac{L-1}{L} \int_{z=0}^{\infty} \log\left(1 + \rho z\right) e^{-z} dz - 2 f_d \log \left(\frac{\frac{\rho}{2 f_d} + 1}{\frac{\rho}{L 2 f_d} + 1} \right).$$
$$(7.34)$$

7.3.2 Optimal Pilot Spacing

Obviously, the lower bound in (7.34) still depends on the pilot spacing L. In case the pilot spacing is not fixed, we can further enhance it by calculating the supremum of (7.34) with respect to L. In this regard, it has to be considered that the pilot spacing L is an integer value. Furthermore, we have to take into account that the derivation of the lower bound in (7.34) is based on the assumption that the pilot spacing is chosen such that the channel fading process is at least sampled with Nyquist rate, i.e., (7.2) has to be fulfilled. In case the pilot spacing L is chosen larger than Nyquist rate, the estimation error process is no longer stationary, which is required for our derivation. At this point it is also important to remark that periodically inserted pilot symbols do not maximize the achievable rate. For the special case of PSK signaling, it is shown in Chapter 10 that the use of a single pilot symbol, i.e., not periodically inserted pilot symbols, is optimal in the sense that it maximizes the achievable rate. However, in the present chapter we restrict to the assumption of periodically inserted pilot symbols with a pilot spacing

---

[2] Recall that a rectangular PSD $S_h(f)$ corresponds to $r_h(l) = \sigma_h^2 \mathrm{sinc}(2 f_d l)$ which is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a raised cosine shape, whose corresponding correlation function is absolutely summable, see Section 2.1 for a discussion.

fulfilling (7.2), which is customary and reasonable as this enables detection and decoding with manageable complexity.

For these conditions, i.e., positive integer values for L fulfilling (7.2), it can be shown that the lower bound $\mathcal{R}_{L,\text{joint}}\big|_{\text{rect}.S_h(f)}$ in (7.34) is maximized for

$$L_{\text{opt}} = \left\lfloor \frac{1}{2f_d} \right\rfloor . \tag{7.35}$$

To prove this statement we differentiate the RHS of (7.34) with respect to L and set the result equal to zero, which yields that the RHS of (7.34) has a unique local extremum at

$$\tilde{L}_{\text{opt}} = \frac{1}{2f_d} \frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)} . \tag{7.36}$$

Numerical evaluation shows that the factor $\frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)}$ is larger than one. As (7.36) is the only local extremum of the RHS of (7.34), the constraint in (7.2), the fact that L is an integer value, and the fact that $\mathcal{R}_{L,\text{joint}}\big|_{\text{rect}.S_h(f)}$ monotonically increases with L for $L < \tilde{L}_{\text{opt}}$ shows that the lower bound is maximized by L_{opt} in (7.35).

Substituting L in (7.34) by L_{opt} in (7.35) yields a lower bound on the achievable rate with joint processing in case the pilot spacing can be arbitrarily chosen while fulfilling (7.2).

7.4 Numerical Evaluation

Fig. 7.1 shows a comparison of the bounds on the achievable rate for synchronized detection based on a solely pilot based channel estimate as discussed in Chapter 5 and the lower bound on the achievable rate with joint processing of data and pilot symbols given in the present chapter.

On the one hand, the lower bound on the achievable rate for joint processing in (7.34) is compared to bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation, i.e, (5.39) and (5.42), for a fixed pilot spacing. As the upper and lower bound on the achievable rate with synchronized detection and a solely pilot based channel estimation are relatively tight, we choose the pilot spacing such that the lower bound on the achievable rate for synchronized detection and a solely pilot based channel estimation in (5.39) is maximized. It can be seen that except for very high channel dynamics, i.e., very large f_d, the lower bound on the achievable rate for joint processing is larger than the bounds on the achievable rate with synchronized detection using a solely pilot based channel estimation. This indicates the possible gain while using joint processing of data and pilot
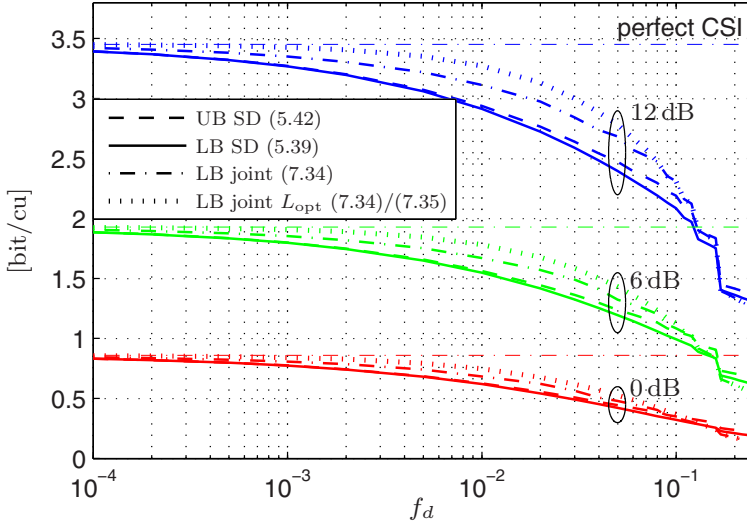
Fig. 7.1: **Comparison of bounds on the achievable rate with synchro-
nized detection and a solely pilot based channel estimation (SD) to
lower bounds on the achievable rate with joint processing of data
and pilot symbols**; except of *LB joint L_{opt}* the pilot spacing L is chosen
such that the lower bound for SD with a solely pilot based channel estimation
(5.39) is maximized; rectangular PSD $S_h(f)$

symbols for a given pilot spacing. Note, the observation that the lower bound
for joint processing for large f_d is smaller than the bounds on the achievable
rate with synchronized detection and a solely pilot based channel estimation
is a result of the lower bounding, i.e., it indicates that the lower bound is not
tight for these parameters.

On the other hand, also the lower bound on the achievable rate with
joint processing and a pilot spacing that maximizes this lower bound, i.e.,
(7.34) in combination with (7.35), is shown. In this case the pilot spacing
is always chosen such that the channel fading process is sampled by the
pilot symbols with Nyquist rate. Obviously, this lower bound is larger than
or equal to the lower bound for joint processing while choosing the pilot
spacing as it is optimal for synchronized detection and a solely pilot based
channel estimation. This behavior arises from the effect that for synchronized
detection and a solely pilot based channel estimation in case of small f_d a
pilot rate is chosen that is higher than the Nyquist rate of the channel fading
process to enhance the channel estimation quality. In case of a joint processing
all symbols are used for channel estimation anyway. Therefore, a pilot rate
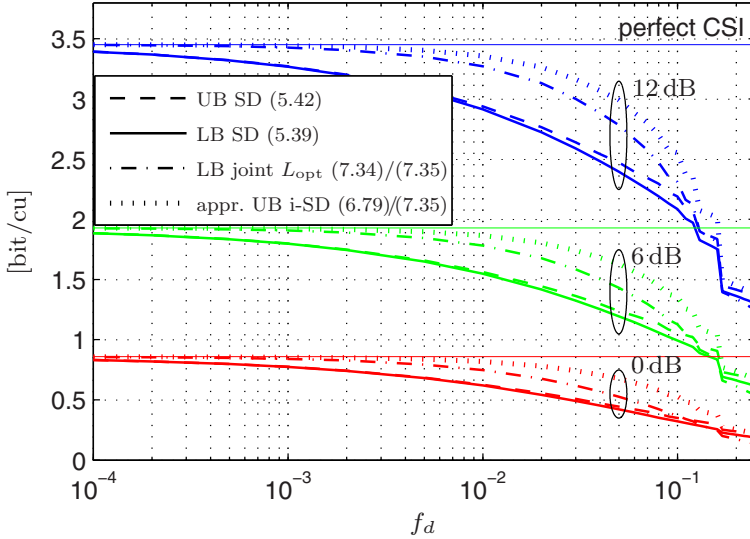
Fig. 7.2: **Comparison of the lower bound on the achievable data rate with joint processing given in (7.34) with the approximate upper bound (6.79) on the achievable rate with the iterative code-aided channel estimation based receiver (i-SD)** described by (6.19), (6.15), and (6.17) ($L = \lfloor 1/(2f_d) \rfloor$ for both, cf. (7.35)); in addition bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation (SD) given in Chapter 5 are shown for a pilot spacing that maximizes the lower bound in (5.39); rectangular PSD $S_h(f)$

higher than Nyquist rate always leads to an increased loss in the achievable rate as less symbols can be used for data transmission.

In Fig. 7.2 the lower bound on the achievable rate with a joint processing of pilot and data symbols given in (7.34) is compared to the approximate upper bound on the achievable rate with the receiver based on iterative code-aided channel estimation (described by (6.19), (6.15), and (6.17)) given in (6.79). For both bounds the pilot spacing has been chosen such that the channel fading process is sampled with Nyquist rate, i.e., $L = \lfloor 1/(2f_d) \rfloor$, see (7.35). Note that the lower bound is an actual lower bound on the achievable rate with joint processing, i.e., with the optimal receiver. In contrast to this, the iterative code-aided channel estimation based receiver described by (6.19), (6.15), and (6.17) is not able to exploit the information corresponding to the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0} | \mathbf{y}_{\backslash 0})$, see Section 6.2.2–6.2.4. Although the given upper bound on the achievable rate is an upper bound for a specific receiver based on iterative code-aided channel estimation, which is not optimal, the upper bound is larger than the lower bound on the achievable rate with joint processing.

In addition, as in Fig. 7.1 for comparison we again show the bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation with a pilot spacing that maximizes the lower bound in (5.39). This comparison gives an indication on the gain that can be expected when using enhanced receivers with code-aided channel estimation in comparison to a solely pilot based channel estimation.
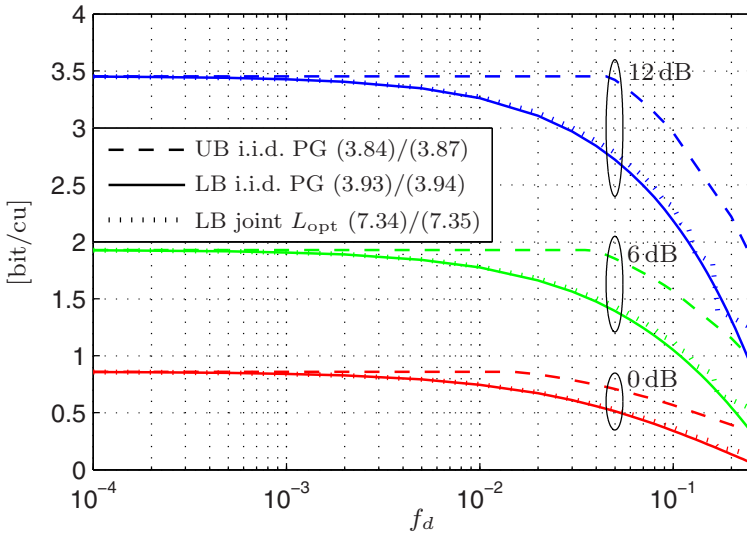

Fig. 7.3: Comparison of the lower bound on the achievable rate with joint processing of data and pilot symbols and a pilot spacing L_{opt} that maximizes this bound, i.e., (7.34) in combination with (7.35), with bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols; rectangular PSD $S_h(f)$

Fig. 7.3 shows the lower bound on the achievable rate for joint processing of data and pilot symbols when choosing L as given in (7.35), which maximizes the lower bound in (7.34). This lower bound is compared to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols that have been given in Chapter 3. Obviously, for some parameters the lower bound on the achievable rate for joint processing of data and pilot symbols, which is achievable with i.i.d. zero-mean proper Gaussian data symbols, is larger than the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., without the assumption of any pilot symbols. However, this observation does not allow to argue that in these cases the use of pilot symbols is better than i.i.d. symbols, as we only compare lower bounds.

7.5 Summary

In the present chapter, we have studied the achievable rate with joint processing of pilot and data symbols in the context of stationary Rayleigh flat-fading channels. Therefor, in the first part we have formally shown that the mutual information between the transmitter and the receiver can be expressed via the solely pilot based channel estimate $\hat{\mathbf{h}}_{\mathrm{pil}}$. This solely pilot based channel estimate is used in the derivation of a lower bound on the achievable rate with joint processing of pilot and data symbol, which is the main contribution of the present chapter.

We have compared this lower bound on the achievable rate with joint processing of pilot and data symbols, on the one hand, with the bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation and, on the other hand, with the upper bound on the achievable rate with the iterative synchronized detection based receiver studied in Chapter 6. These comparisons give an indication on the possible gain in terms of the achievable rate when using a joint processing of pilot and data symbols in comparison to the typically used synchronized detection based on a solely pilot based channel estimation. Furthermore, we have compared the lower bound on the achievable rate with joint processing of pilot and data symbols with the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols derived in Chapter 3.

Chapter 8
MIMO Flat-Fading Channels

In the present chapter, we extend the bounds on the achievable rate with i.d. input symbols given in Chapter 3 for the discrete-time flat-fading SISO channel to the MIMO case, including spatial antenna correlation. Beside the constraint on i.d. input symbols in temporal domain, as it has been also used in the SISO case, we furthermore do not optimize over the spatial input covariance matrix, but restrict to zero-mean input symbols on the individual transmit antennas which are characterized by a spatial covariance as it is optimal in case of a coherent channel.

Within this chapter, we first introduce the MIMO system model. Afterwards, we extend the bounds on the achievable rate given in Chapter 3 to the MIMO scenario. Finally, we compare these bounds to existing bounds on the achievable rate using synchronized detection in combination with a solely pilot based channel estimation.

Like in the SISO case, the main novelty in the present chapter is the derivation of an upper bound on the achievable rate, which is not explicitly based on a peak power constraint. Therefore, it allows to give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. The main derivation steps for the extension of the bounds from the SISO case to the MIMO case are similar to the ones used in [102] where capacity bounds for peak power constrained MIMO wideband channels have been derived.

As the MIMO channel capacity is not the main subject of the present work, for further reading we give a brief overview on literature regarding the MIMO fading channel capacity. A general overview of results on the capacity of MIMO channels is given in [42]. The basic papers concerning the use of multiple antennas and the capacity of coherent spatially uncorrelated MIMO fading channels are [36] and [120]. Results concerning the capacity of spatially correlated MIMO fading channels in the coherent case are, e.g., given in [123], [124], [17], and [57]. Regarding further reading on capacity results for the noncoherent capacity we refer back to Section 1.1. In addition, the scenario of spatial antenna correlation in the noncoherent case has been examined in [53], [114] and [132], where the first and the second one discuss

this problem in the context of a block fading channel and the third one assumes a temporally uncorrelated channel, i.e., both setups are different to ours.

8.1 MIMO System Model

In this section, we extend the SISO system model given in Chapter 2 to the MIMO case.

The input-output relation of the MIMO channel with n_T transmit antennas and n_R receive antennas for an individual time instant k is given by

$$\tilde{\mathbf{y}}(k) = \tilde{\mathbf{H}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k) \tag{8.1}$$

where $\tilde{\mathbf{x}}(k) = [x_1(k)\ldots x_{n_T}(k)]^T \in \mathbb{C}^{n_T \times 1}$ is the channel input vector transmitted over the n_T transmit antennas at time k, the $n_R \times n_T$ matrix $\tilde{\mathbf{H}}(k)$ is defined by

$$\left[\tilde{\mathbf{H}}(k)\right]_{ij} = h_{ij}(k) \tag{8.2}$$

and represents the channel fading matrix with the complex-valued fading coefficients $h_{ij}(k), \quad 1 \le i \le n_R, \quad 1 \le j \le n_T$ of the different subchannels. Furthermore, $\tilde{\mathbf{n}}(k) = [n_1(k)\ldots n_{n_R}(k)]^T \in \mathbb{C}^{n_R \times 1}$ is the additive white Gaussian noise vector. $\tilde{\mathbf{y}}(k) = [y_1(k)\ldots y_{n_R}(k)]^T \in \mathbb{C}^{n_R \times 1}$ is the received signal vector at time instant k.

Corresponding to the SISO case, the study of the achievable rate is based on a matrix-vector notation incorporating the temporal domain. Therefore, we use the following representation, considering a transmission duration of N symbol intervals

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{8.3}$$

where $\mathbf{h}$ results from stacking the vectors

$$\bar{\mathbf{h}}_{ij} = [h_{ij}(1)\ldots h_{ij}(N)]^T \in \mathbb{C}^{N \times 1}, \quad 1 \le i \le n_R, 1 \le j \le n_T \tag{8.4}$$

containing the fading weights of the subchannel from transmit antenna j to receive antenna i over time, as follows

$$\mathbf{h} = \left[\bar{\mathbf{h}}_{11}^T \ldots \bar{\mathbf{h}}_{1n_T}^T \ldots \bar{\mathbf{h}}_{n_R 1}^T \ldots \bar{\mathbf{h}}_{n_R n_T}^T\right]^T \in \mathbb{C}^{N \cdot n_T \cdot n_R \times 1}. \tag{8.5}$$

Define

$$\bar{\mathbf{x}}_j = [x_j(1)\ldots x_j(N)]^T \in \mathbb{C}^{N \times 1}, \quad 1 \le j \le n_T \tag{8.6}$$
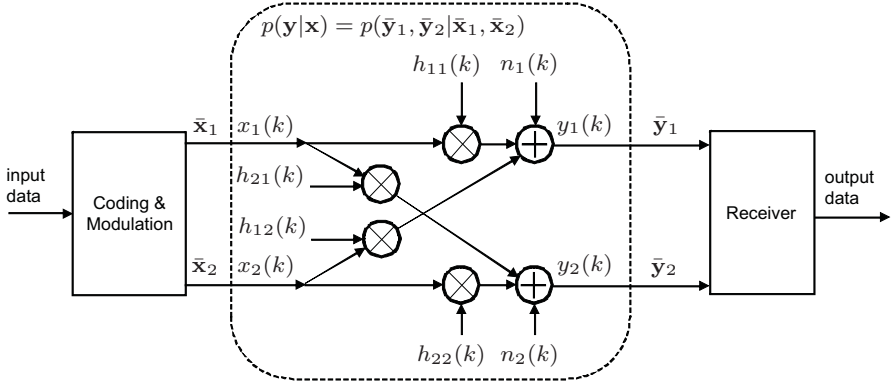
Fig. 8.1: Simplified block diagram of transmission system based on a discrete-time flat-fading MIMO channel with $n_T = n_R = 2$ antennas

to be the vector consisting of the transmit symbols of transmit antenna j over time, and $\bar{\mathbf{X}}_j = \mathrm{diag}(\bar{\mathbf{x}}_j)$. Hence, the $N \cdot n_R \times N \cdot n_T \cdot n_R$ matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \mathbf{I}_{n_R} \otimes \left[\bar{\mathbf{X}}_1 \ldots \bar{\mathbf{X}}_{n_T} \right] \tag{8.7}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{I}_{n_R}$ the $n_R \times n_R$ identity matrix. Furthermore, the received signal vector $\mathbf{y}$ and the noise vector $\mathbf{n}$ are given by $\mathbf{y} = \left[\bar{\mathbf{y}}_1^T \ldots \bar{\mathbf{y}}_{n_R}^T \right]^T \in \mathbb{C}^{N \cdot n_R \times 1}$ and $\mathbf{n} = \left[\bar{\mathbf{n}}_1^T \ldots \bar{\mathbf{n}}_{n_R}^T \right]^T \in \mathbb{C}^{N \cdot n_R \times 1}$ where

$$\bar{\mathbf{y}}_i = [y_i(1) \ldots y_i(N)]^T \in \mathbb{C}^{N \times 1}, \qquad 1 \leq i \leq n_R \tag{8.8}$$

$$\bar{\mathbf{n}}_i = [n_i(1) \ldots n_i(N)]^T \in \mathbb{C}^{N \times 1}, \qquad 1 \leq i \leq n_R. \tag{8.9}$$

Analogously we define $\mathbf{x} = \left[\bar{\mathbf{x}}_1^T \ldots \bar{\mathbf{x}}_{n_T}^T \right]^T$. Fig. 8.1 shows exemplarily a block-diagram of a 2×2 MIMO system.

In the following, we describe the statistical properties of the different quantities. First, we assume the processes $\{\tilde{\mathbf{H}}(k)\}$, $\{\tilde{\mathbf{x}}(k)\}$, and $\{\tilde{\mathbf{n}}(k)\}$ to be mutually independent.

Furthermore, we assume that the elements of the noise sequence $\{n_j(k)\}$ at receive antenna j are i.i.d. proper Gaussian random variables of zero-mean and variance σ_n^2. In addition, the noise sequences at the different receive antennas are mutually uncorrelated, so that

$$\mathbf{R_n} = \mathrm{E}[\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I}_{N \cdot n_R}. \tag{8.10}$$

Concerning the channel input, we assume that the transmit sequences at the individual transmit antennas $\{x_j(k)\}$ consist of zero-mean i.d. input symbols. Furthermore, we assume that the covariance of the input symbols

over the different transmit antennas at time instant k is given by

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \mathrm{E}\left[\tilde{\mathbf{x}}(k)(\tilde{\mathbf{x}}(k))^H\right], \quad \forall k. \tag{8.11}$$

The average transmit power is limited to P, i.e.,

$$P \geq \mathrm{E}\left[(\tilde{\mathbf{x}}(k))^H \tilde{\mathbf{x}}(k)\right]. \tag{8.12}$$

Notice that in case of a non-peak power constrained input it is optimal to use the maximum average transmit power, see the discussion in Section 3.4.1.1.

For the MIMO system, we define the nominal SNR ρ as

$$\rho = \frac{P\sigma_h^2}{\sigma_n^2}. \tag{8.13}$$

We name ρ the *nominal* mean SNR, as it only corresponds to the mean SNR if the maximum average power P is used, i.e., if (8.12) holds with equality. As for non-peak power constrained input symbols the achievable rate is maximized by using the maximum average transmit power P, in this case ρ corresponds to the actual mean SNR.

As in the SISO case, the channel fading processes $\{h_{ij}(k)\}$ are zero-mean jointly proper Gaussian. Concerning their temporal statistics, we make the same assumptions as in the SISO case, see Chapter 2. Therefore, the temporal correlation of each individual subchannel process $\{h_{ij}(k)\}$ is given by (2.2) and the PSD (2.3), where we assume that all subchannels are characterized by the same PSD. Therefore, the temporal correlation matrix of the individual subchannels is given by

$$\mathbf{R}_{\bar{h}} = \mathrm{E}\left[\bar{\mathbf{h}}_{ij}\bar{\mathbf{h}}_{ij}^H\right], \quad \forall i, j. \tag{8.14}$$

8.1.1 Spatially Uncorrelated Channel

In case of spatially uncorrelated transmit antennas, we assume uncorrelated input symbols at the different transmit antennas, which are capacity-achieving in case of perfect channel knowledge, i.e., $\mathbf{R}_{\tilde{\mathbf{x}}}$ is diagonal, see [120]. Their maximum average individual transmit power is given as

$$\sigma_x^2 = \frac{P}{n_T}. \tag{8.15}$$

As in case of a spatially uncorrelated channel the fading processes $\{h_{ij}(k)\}$ of the different subchannels are independent, we get

$$\mathbf{R_h} = \mathrm{E}\left[\mathbf{h}\mathbf{h}^H\right] = \mathbf{I}_{n_R} \otimes \mathbf{I}_{n_T} \otimes \mathbf{R}_{\bar{h}}. \tag{8.16}$$

8.1.2 Spatial Antenna Correlation

We also want to discuss the achievable rate of MIMO systems with spatial correlation corresponding to the *separable correlation model*, see, e.g., [109] and [123]. This model falls into the class of the unitary independent unitary (UIU) model [123]. We now show how this model can be attributed to the model of a MIMO system without spatial correlation using a unitary transformation. The same approach has also been used in [102] and [53].

We assume that the spatial channel correlation of the transmit and the receive antennas are separable, allowing the following separation of the channel matrix

$$\tilde{\mathbf{H}}(k) = \tilde{\mathbf{\Theta}}_R^{\frac{1}{2}} \tilde{\mathbf{H}}_w(k) \left(\tilde{\mathbf{\Theta}}_T^{\frac{1}{2}} \right)^T \tag{8.17}$$

where $\tilde{\mathbf{\Theta}}_T$ and $\tilde{\mathbf{\Theta}}_R$ are the spatial transmit and receive correlation matrices, being independent of the time index k and deterministic. The elements of the matrix $\tilde{\mathbf{H}}_w(k)$ are i.i.d. zero-mean proper Gaussian with variance σ_h^2.

The eigenvalue decomposition of the spatial transmit and receive correlation matrices are given by

$$\tilde{\mathbf{\Theta}}_T = \tilde{\mathbf{U}}_T \tilde{\mathbf{\Phi}} \tilde{\mathbf{U}}_T^H \tag{8.18}$$

$$\tilde{\mathbf{\Theta}}_R = \tilde{\mathbf{U}}_R \tilde{\mathbf{\Psi}} \tilde{\mathbf{U}}_R^H \tag{8.19}$$

where the matrices $\tilde{\mathbf{U}}_T$ and $\tilde{\mathbf{U}}_R$ are unitary and their columns are the eigenvectors of $\tilde{\mathbf{\Theta}}_T$ and $\tilde{\mathbf{\Theta}}_R$. Furthermore, the matrices $\tilde{\mathbf{\Phi}} = \text{diag}(\phi_1, \ldots, \phi_{n_T})$ and $\tilde{\mathbf{\Psi}} = \text{diag}(\psi_1, \ldots, \psi_{n_R})$ are diagonal and contain the eigenvalues of the spatial transmit and receive antenna correlation. Using these decompositions, the system model in (8.1) can be transformed as follows

$$\tilde{\mathbf{y}}(k) = \tilde{\mathbf{H}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k)$$
$$= \tilde{\mathbf{\Theta}}_R^{\frac{1}{2}} \tilde{\mathbf{H}}_w(k) \left(\tilde{\mathbf{\Theta}}_T^{\frac{1}{2}} \right)^T \tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k)$$
$$= \tilde{\mathbf{U}}_R \tilde{\mathbf{\Psi}}^{\frac{1}{2}} \tilde{\mathbf{U}}_R^H \tilde{\mathbf{H}}_w(k) \tilde{\mathbf{U}}_T^* \tilde{\mathbf{\Phi}}^{\frac{1}{2}} \tilde{\mathbf{U}}_T^T \tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k). \tag{8.20}$$

In the following, we use the unitary transformation given in [53], [102] to get a spatially decorrelated system model

$$\tilde{\mathbf{U}}_R^H \tilde{\mathbf{y}}(k) = \tilde{\mathbf{\Psi}}^{\frac{1}{2}} \tilde{\mathbf{U}}_R^H \tilde{\mathbf{H}}_w(k) \tilde{\mathbf{U}}_T^* \tilde{\mathbf{\Phi}}^{\frac{1}{2}} \tilde{\mathbf{U}}_T^T \tilde{\mathbf{x}}(k) + \tilde{\mathbf{U}}_R^H \tilde{\mathbf{n}}(k)$$
$$\Leftrightarrow \breve{\mathbf{y}}(k) = \tilde{\mathbf{\Psi}}^{\frac{1}{2}} \breve{\mathbf{H}}_w(k) \tilde{\mathbf{\Phi}}^{\frac{1}{2}} \breve{\mathbf{x}}(k) + \breve{\mathbf{n}}(k) \tag{8.21}$$

where

$$\check{\mathbf{y}}(k) = \tilde{\mathbf{U}}_R^H \tilde{\mathbf{y}}(k) \tag{8.22}$$

$$\check{\mathbf{n}}(k) = \tilde{\mathbf{U}}_R^H \tilde{\mathbf{n}}(k) \tag{8.23}$$

$$\check{\mathbf{H}}_w(k) = \tilde{\mathbf{U}}_R^H \tilde{\mathbf{H}}_w(k) \tilde{\mathbf{U}}_T^* \tag{8.24}$$

$$\check{\tilde{\mathbf{x}}}(k) = \tilde{\mathbf{U}}_T^T \tilde{\mathbf{x}}(k). \tag{8.25}$$

Due to the independence of the elements of $\tilde{\mathbf{n}}(k)$ and the unitarity of $\tilde{\mathbf{U}}_R^H$, $\check{\mathbf{n}}(k)$ and $\tilde{\mathbf{n}}(k)$ are identically distributed, i.e., $\check{\mathbf{n}}(k) \sim \tilde{\mathbf{n}}(k)$. Using the same argumentation, we get $\check{\mathbf{H}}_w(k) \sim \tilde{\mathbf{H}}_w(k)$. Furthermore, unitary transformations preserve norms, therefore $\check{\tilde{\mathbf{x}}}(k)$ fulfills the same power constraint as $\tilde{\mathbf{x}}(k)$. Finally, the transformed channel output vector $\check{\mathbf{y}}(k)$ is a sufficient statistic of $\tilde{\mathbf{y}}(k)$. Due to these properties, the spatially decorrelated system model in (8.21) exhibits the same mutual information as the original one in (8.1). The elements of the transmit vector $\check{\tilde{\mathbf{x}}}(k)$ contain the signal transmitted on the different transmit eigenmodes and do not directly correspond to the symbols transmitted over the individual antennas. The same holds for the channel output vector $\check{\mathbf{y}}(k)$, whose elements represent the observations on the individual receive eigenmodes.

8.1.2.1 Spatial Channel Input Correlation

We would still have to find the spatial covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$ of the input symbols, see (8.11), which maximizes the achievable rate. As this is out of the scope of the present work, we apply the spatial input covariance that is optimal for the case of perfect CSI at the receiver and statistical channel knowledge at the transmitter. Furthermore, we guess that for the important case of small channel dynamics the input distribution that is optimal for the coherent case will not be highly non-optimal for the noncoherent case.

Let the eigenvalue decomposition of the spatial transmit signal correlation matrix be given by

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \mathbf{V}\alpha\mathbf{P}\mathbf{V}^H \tag{8.26}$$

where the columns of the unitary matrix $\mathbf{V}$ are the eigenvectors of $\mathbf{R}_{\tilde{\mathbf{x}}}$ and the diagonal matrix $\alpha\mathbf{P} = \alpha\mathrm{diag}(p_1, \ldots, p_{n_T})$ contains the eigenvalues of $\mathbf{R}_{\tilde{\mathbf{x}}}$ which correspond to the average transmit powers put onto the transmit eigenmodes. The factor $\alpha \in [0, 1]$ allows for average transmit powers being smaller than the maximum average transmit power P. Hence,

$$\sum_{j=1}^{n_T} p_j = P. \tag{8.27}$$

For perfect channel state information at receiver side, in [57] and [123] it has been shown that the capacity-achieving input correlation is character-

ized by eigenvectors corresponding to the eigenvectors of the spatial channel transmit covariance matrix, i.e.,

$$\mathbf{V} = \tilde{\mathbf{U}}_T^*. \tag{8.28}$$

Introducing this into the spatially decorrelated channel model in (8.21), we get

$$\breve{\mathbf{y}}(k) = \tilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\breve{\mathbf{H}}_w(k)\tilde{\boldsymbol{\Phi}}^{\frac{1}{2}}\tilde{\mathbf{P}}^{\frac{1}{2}}\breve{\mathbf{x}}(k) + \breve{\mathbf{n}}(k) \tag{8.29}$$

with

$$\tilde{\mathbf{P}} = \frac{\mathbf{P}}{\sigma_x^2} \tag{8.30}$$

and $\breve{\mathbf{x}}(k)$ being a vector with i.i.d. zero-mean symbols with variance $\alpha\sigma_x^2$ and $\sigma_x^2 = P/n_T$. For the coherent and non-peak power constrained case, the optimum transmit powers p_j have been numerically evaluated in [124] by iteratively solving an implicit system of equations. A closed form solution for the optimum transmit powers is as far as we know not available.

   Remark: To keep notation simple, in the following derivation we use (8.29) instead of (8.1) without changing the notation of the temporally stacked representation in (8.3). Comparing (8.29) with (8.1), and identifying the $\tilde{\mathbf{H}}$ with $\tilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\breve{\mathbf{H}}_w(k)\tilde{\boldsymbol{\Phi}}^{\frac{1}{2}}\tilde{\mathbf{P}}^{\frac{1}{2}}$, the correlation matrix of the stacked channel vector $\mathbf{h}$ in (8.5), including the transmit power allocation $\tilde{\mathbf{P}}$, is given by

$$\mathbf{R_h} = \mathrm{E}\left[\mathbf{h}\mathbf{h}^H\right] = \tilde{\boldsymbol{\Psi}} \otimes (\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}}) \otimes \mathbf{R_{\bar{h}}}. \tag{8.31}$$

Thus, for the evaluation of a spatially correlated system in comparison to a spatially uncorrelated system, (8.16) has to be substituted by (8.31) for $\mathbf{R_h}$.

8.2 Bounds on the Achievable Rate

Corresponding to the SISO case in Chapter 3, we study the achievable rate of a MIMO system with i.d. zero-mean input symbols in the temporal domain. Analogous to the SISO case the restriction to i.d. input symbols in temporal domain is required in the following derivations for mathematical reasons. Furthermore, we assume the input symbols on the individual transmit antennas to be zero-mean and we assume the spatial covariance matrix of the transmit symbols on the individual transmit antennas to be given by the one that is capacity-achieving in the coherent case. On the one hand, we give bounds on the achievable rate for the case of a peak power constraint and on the other hand for the special case of i.i.d. zero-mean proper Gaussian input symbols.

Corresponding to (3.1) the ergodic capacity C of the MIMO channel is given by

$$C = \sup_{\mathcal{P}_{\mathrm{MIMO}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \qquad (8.32)$$

where $\mathcal{P}_{\mathrm{MIMO}}$ is the set containing all PDFs $p(\mathbf{x})$ over the complex set $\mathbb{C}^{N \cdot n_T}$ with an average power P, i.e.,

$$\mathcal{P}_{\mathrm{MIMO}} = \left\{ p(\mathbf{x}) \big| \mathbf{x} \in \mathbb{C}^{N \cdot n_T}, \frac{1}{N} \mathrm{E} \left[\mathbf{x}^H \mathbf{x} \right] \leq P \right\}. \qquad (8.33)$$

As in Chapter 3, we simplify the problem and do not study the capacity in general but restrict to i.d. input symbols $\check{\mathbf{x}}(k)$, see (8.29), i.e., the vector input symbols at the different time instances are i.d., as it has already been stated in Section 8.1. In addition, we restrict to zero-mean input symbols $\check{\mathbf{x}}(k)$. Furthermore, we assume the spatial transmit covariance matrix $\mathbf{R}_{\check{\mathbf{x}}}$ to be fixed to the one that is capacity-achieving in the coherent case as it has been discussed in Section 8.1.1 and Section 8.1.2.1. As we discuss the spatially correlated case based on the unitary transformation in (8.29), and use the simplified notation as discussed in the remark at the end of Section 8.1.2.1, in the following we assume that

$$\mathbf{R}_{\check{\mathbf{x}}} = \alpha \sigma_x^2 \mathbf{I}_{n_T} = \alpha \frac{P}{n_T} \mathbf{I}_{n_T} \qquad (8.34)$$

where the factor $\alpha \in [0, 1]$ allows for using an average transmit power being smaller than the maximum average transmit power P, cf.(8.12).

Furthermore, the normalized input power matrix $\tilde{\mathbf{P}}$ in (8.30) is chosen such that it maximizes the coherent capacity, see Section 8.1.2.1 under the constraint trace $\left[\tilde{\mathbf{P}} \right] = n_T$.

Therefore, in the following we use the term *achievable rate* R which we define as

$$R = \sup_{\mathcal{P}_{\mathrm{MIMO, \, i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \qquad (8.35)$$

with the set $\mathcal{P}_{\mathrm{MIMO, \, i.d.}}$ given by

$$\mathcal{P}_{\mathrm{MIMO, \, i.d.}} = \Bigg\{ p(\mathbf{x}) \bigg| \mathbf{x} \in \mathbb{C}^{N \cdot n_T}, \quad p(\check{\mathbf{x}}(k)) = p(\check{\mathbf{x}}(l)) \quad \forall k, l, \quad \tilde{\mathbf{P}} \text{ fixed}$$

$$\left\{ \mathrm{E}\left[\check{\mathbf{x}}(k) \right] = 0, \ \mathrm{E}\left[\check{\mathbf{x}}(k) \left(\check{\mathbf{x}}(k) \right)^H \right] = \frac{\alpha P}{n_T} \mathbf{I}_{n_T}, \alpha \in [0, 1], \forall k \right\} \Bigg\}$$

$$(8.36)$$

where we have used the substitution of (8.1) by (8.29), see remark at the end of Section 8.1.2.1.

At some specific points, we further introduce a peak power constraint on the input symbols. Therefore, we define the following set of input distributions

$$
\mathcal{P}_{\mathrm{MIMO,\ i.d.}}^{\mathrm{peak}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^{N \cdot n_T}, \quad p(\check{\mathbf{x}}(k)) = p(\check{\mathbf{x}}(l)) \quad \forall k, l, \quad \tilde{\mathbf{P}} \text{ fixed,} \right.
$$

$$
\left\{ \mathrm{E}\left[\check{\mathbf{x}}(k)\right] = 0, \mathrm{E}\left[\check{\mathbf{x}}(k)\left(\check{\mathbf{x}}(k)\right)^H\right] = \frac{\alpha P}{n_T} \mathbf{I}_{n_T}, \alpha \in [0, 1], \right.
$$

$$
\left. p(\check{\mathbf{x}}(k)) = 0 \ \forall \check{\mathbf{x}}(k) : (\check{\mathbf{x}}(k))^H \check{\mathbf{x}}(k) > P_{\mathrm{peak}} \right\} \forall k \left. \right\} \quad (8.37)
$$

which corresponds to the set $\mathcal{P}_{\mathrm{MIMO,\ i.d.}}$ with a further restriction to the peak power, i.e., $(\check{\mathbf{x}}(k))^H \check{\mathbf{x}}(k) \leq P_{\mathrm{peak}}$.

As in Chapter 3, the derivation is based on the separation of the mutual information rate given in (3.10). Therefore, in Section 8.2.1 the lower and the upper bound on the channel output entropy rate $h'(\mathbf{y})$ derived in Section 3.2.1 and Section 3.2.2 are extended to the MIMO case. Afterwards, in Section 8.2.2 the upper and lower bound on the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ derived in Section 3.3 for the SISO case are extended to the MIMO case.

8.2.1 The Received Signal Entropy Rate $h'(\mathbf{y})$

8.2.1.1 Lower Bound on $h'(\mathbf{y})$

Corresponding to the derivation in Section 3.2.1 the channel output entropy rate is lower-bounded by

$$
h'(\mathbf{y}) \geq h'_L(\mathbf{y}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}). \quad (8.38)
$$

The first term on the RHS of (8.38) $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is the mutual information rate in case of perfect channel knowledge. As it depends on the input distribution, we postpone its evaluation to Section 8.2.3.2 where we derive bounds on the achievable rate.

The second term on the RHS of (8.38) originates from AWGN and, thus, can be calculated as

$$
h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) = n_R \cdot \log \left(\pi e \sigma_n^2 \right). \quad (8.39)
$$

Introducing (8.39) into (8.38) gives a lower bound on $h'(\mathbf{y})$.

8.2.1.2 Upper Bound on $h'(\mathbf{y})$

Analogously to the SISO case, we make use of the fact that the entropy $h(\mathbf{y})$ of the complex random vector $\mathbf{y}$ of dimension $N \cdot n_R$ with nonsingular correlation matrix $\mathbf{R}_y = \mathrm{E}[\mathbf{y}\mathbf{y}^H]$ is upper-bounded by [82]

$$h(\mathbf{y}) \leq \log\left[(\pi e)^{N \cdot n_R} \det(\mathbf{R_y})\right]. \tag{8.40}$$

Based on the spatially decorrelated input-output relation in (8.29) the correlation of the channel output $\mathbf{\check{y}}$ is given by

$$
\begin{aligned}
\mathbf{R_{\check{y}}} &= \mathrm{E}_{\check{\mathbf{y}}}\left[\check{\mathbf{y}}\check{\mathbf{y}}^H\right] \\
&= \mathrm{E}\left[\tilde{\mathbf{\Psi}}^{\frac{1}{2}}\check{\mathbf{H}}_w(k)\tilde{\mathbf{\Phi}}^{\frac{1}{2}}\tilde{\mathbf{P}}^{\frac{1}{2}}\check{\mathbf{x}}(k)(\check{\mathbf{x}}(k))^H\tilde{\mathbf{P}}^{\frac{1}{2}}\tilde{\mathbf{\Phi}}^{\frac{1}{2}}(\check{\mathbf{H}}_w(k))^H\tilde{\mathbf{\Psi}}^{\frac{1}{2}}\right] + \sigma_n^2\mathbf{I}_{n_R} \\
&= \tilde{\mathbf{\Psi}}\sigma_h^2\alpha\sum_{j=1}^{n_T}(p_j\phi_j) + \sigma_n^2\mathbf{I}_{n_R}.
\end{aligned}
\tag{8.41}
$$

The factor $\alpha \in [0,1]$ allows for an average transmit power smaller than P. Consequently, the channel output entropy rate can be upper-bounded by

$$
\begin{aligned}
h'(\mathbf{y}) &\leq \frac{1}{N}\log\left[(\pi e)^{N \cdot n_R}\det(\mathbf{R_y})\right] \\
&\overset{(a)}{\leq} \log\left[(\pi e)^{n_R}\det(\mathbf{R_{\check{y}}})\right] \\
&= n_R\log(\pi e) + \sum_{i=1}^{n_R}\log\left(\psi_i\sigma_h^2\alpha\sum_{j=1}^{n_T}(p_j\phi_j) + \sigma_n^2\right) = h'_U(\mathbf{y}).
\end{aligned}
\tag{8.42}
$$

where (a) results from discarding the temporal correlation.

For the special case of a spatially uncorrelated channel we get

$$h'(\mathbf{y}) \leq n_R\log\left(\pi e\left(\alpha P\sigma_h^2 + \sigma_n^2\right)\right) = h'_U(\mathbf{y}). \tag{8.43}$$

8.2.2 The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$

In this section, we extend the upper and the lower bound on $h'(\mathbf{y}|\mathbf{x})$, given in Section 3.3 for the SISO case, to the MIMO case.

8.2.2.1 Upper Bound on $h'(\mathbf{y}|\mathbf{x})$

Corresponding to the SISO case the probability density function of $\mathbf{y}$ conditioned on $\mathbf{x}$ is zero-mean proper Gaussian. Its covariance matrix $\mathbf{R_{y|x}}$ is

given by

$$\mathbf{R_{y|x}} = \mathbf{X}\mathbf{R_h}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_{N \cdot n_R}$$

$$= \mathbf{X} \left(\tilde{\boldsymbol{\Psi}} \otimes (\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}}) \otimes \mathbf{R_{\bar{h}}} \right) \mathbf{X}^H + \sigma_n^2 \mathbf{I}_{N \cdot n_R} \tag{8.44}$$

where we used (8.31). Thus, the entropy $h(\mathbf{y}|\mathbf{x})$ is given by

$$h(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{\mathbf{x}} \left[\log \left((\pi e)^{N \cdot n_R} \det(\mathbf{R_{y|x}}) \right) \right] \tag{8.45}$$

$$= \mathrm{E}_{\mathbf{x}} \left[\sum_{i=1}^{n_R} \log \left((\pi e)^N \det \left(\psi_i \sum_{j=1}^{n_T} \frac{p_j}{\sigma_x^2} \phi_j \bar{\mathbf{X}}_j \mathbf{R_{\bar{h}}} \bar{\mathbf{X}}_j^H + \sigma_n^2 \mathbf{I}_N \right) \right) \right]. \tag{8.46}$$

In the next steps, we principally follow the same ideas as they have been used in [102]. By rewriting the sum of matrices in (8.46) by means of stacked matrices we get

$$h(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{\mathbf{x}} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \left(\bar{\mathbf{X}}_1 \dots \bar{\mathbf{X}}_{n_T} \right) \left((\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}}) \otimes \mathbf{R_{\bar{h}}} \right) \begin{pmatrix} \bar{\mathbf{X}}_1^H \\ \vdots \\ \bar{\mathbf{X}}_{n_T}^H \end{pmatrix} + \mathbf{I}_N \right) \right]$$

$$+ Nn_R \log \left(\pi e \sigma_n^2 \right)$$

$$\stackrel{(a)}{=} \mathrm{E}_{\mathbf{x}} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \begin{pmatrix} \bar{\mathbf{X}}_1^H \\ \vdots \\ \bar{\mathbf{X}}_{n_T}^H \end{pmatrix} \left(\bar{\mathbf{X}}_1 \dots \bar{\mathbf{X}}_{n_T} \right) \left((\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}}) \otimes \mathbf{R_{\bar{h}}} \right) + \mathbf{I}_{Nn_T} \right) \right]$$

$$+ Nn_R \log \left(\pi e \sigma_n^2 \right)$$

$$\stackrel{(b)}{\leq} \sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i \alpha \sigma_x^2}{\sigma_n^2} (\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}}) \otimes \mathbf{R_{\bar{h}}} + \mathbf{I}_{N \cdot n_T} \right) + Nn_R \log \left(\pi e \sigma_n^2 \right)$$

$$\stackrel{(c)}{=} \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \log \det \left(\frac{\psi_i \alpha \phi_j p_j}{\sigma_n^2} \mathbf{R_{\bar{h}}} + \mathbf{I}_N \right) + Nn_R \log \left(\pi e \sigma_n^2 \right)$$

$$\stackrel{(d)}{=} \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \log \det \left(\frac{\psi_i \alpha \phi_j p_j}{\sigma_n^2} \mathbf{U}\boldsymbol{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) + Nn_R \log \left(\pi e \sigma_n^2 \right)$$

$$= \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \sum_{k=1}^{N} \log \left(\frac{\psi_i \alpha \phi_j p_j}{\sigma_n^2} \lambda_k + 1 \right) + Nn_R \log \left(\pi e \sigma_n^2 \right). \tag{8.47}$$

Here, (a) follows from the fact that for a matrix $\mathbf{A}$ of dimension $m \times n$ and a matrix $\mathbf{B}$ of dimension $n \times m$ with $m \leq n$, the product $\mathbf{B}\mathbf{A}$ has the same eigenvalues as $\mathbf{A}\mathbf{B}$, counting multiplicity, together with an additional $n - m$ eigenvalues equal to 0, [52, Theorem 1.3.20]. Equation (3.25) is a special case of this theorem. For (b), we have used the concavity of log det on

the set of positive definite matrices. Furthermore, we assume zero-mean input symbols on the individual transmit antennas and have used the independency of the elements of $\bar{\mathbf{x}}_k$ and $\bar{\mathbf{x}}_l$ for $k \neq l$, i.e., the independency of the transmit symbols on the individual transmit eigenmodes, which for the special case of a spatially uncorrelated channel corresponds to the independency of the symbols transmitted over the individual transmit antennas. In addition, we have used that the average power per eigenmode $\mathrm{E}\left[[\breve{\mathbf{x}}(k)]_l\right] \quad \forall k,l$, i.e., before weighting with the matrix $\tilde{\mathbf{P}}$, see (8.29), is given by $\alpha\sigma_x^2$. Equality (c) follows from the the fact that the matrix $(\tilde{\boldsymbol{\Phi}}\tilde{\mathbf{P}})$ is diagonal. For (d), we have used the spectral decomposition of the channel correlation matrix $\mathbf{R}_{\bar{\mathbf{h}}}$ as in the SISO case, see (3.21).

As in the SISO case, we evaluate the case $N \to \infty$ based on Szegö's theorem and finally get the following upper bound

$$
\begin{aligned}
h'(\mathbf{y}|\mathbf{x}) &\leq h'_U(\mathbf{y}|\mathbf{x}) \\
&= \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i \alpha p_j \phi_j}{\sigma_n^2} S_h(f) + 1\right) df + n_R \log\left(\pi e \sigma_n^2\right). \quad (8.48)
\end{aligned}
$$

For the special case of a spatially uncorrelated channel, (8.48) simplifies to

$$
h'_U(\mathbf{y}|\mathbf{x}) = n_R n_T \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\alpha P}{n_T \sigma_n^2} S_h(f) + 1\right) df + n_R \log\left(\pi e \sigma_n^2\right). \quad (8.49)
$$

Obviously, the pre-log of the first summand of this bound scales with the number of transmit antennas n_T and the number of receive antennas n_R and the support of $S_h(f)$.

8.2.2.2 Lower Bound on $h'(\mathbf{y}|\mathbf{x})$ for a Rectangular PSD

In this section, we extend the lower bound on $h'(\mathbf{y}|\mathbf{x})$ for a rectangular PSD, which is derived in Section 3.3.2 for the SISO case, to the MIMO scenario.

The entropy $h(\mathbf{y}|\mathbf{x})$ given in (8.46) can be lower-bounded as follows

$$
\begin{aligned}
h(\mathbf{y}|\mathbf{x}) &= \mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{n_R} \log\left(\det\left(\frac{\psi_i}{\sigma_n^2}\left(\sum_{j=1}^{n_T} \frac{p_j}{\sigma_x^2}\phi_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j^H\right) \odot \mathbf{R}_{\bar{\mathbf{h}}} + \mathbf{I}_N\right)\right)\right] \\
&\quad + N n_R \log\left(\pi e \sigma_n^2\right) \\
&\overset{(a)}{\geq} \mathrm{E}_{\mathbf{x}}\left[\sum_{i=1}^{n_R} \log\left(\det\left(\frac{\psi_i}{\sigma_n^2}\left(\mathbf{I}_N \odot \sum_{j=1}^{n_T} \frac{p_j}{\sigma_x^2}\phi_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j^H\right) \mathbf{R}_{\bar{\mathbf{h}}} + \mathbf{I}_N\right)\right)\right] \\
&\quad + N n_R \log\left(\pi e \sigma_n^2\right)
\end{aligned}
$$

$$= \mathrm{E_x} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \mathbf{Z} \mathbf{R}_{\bar{\mathbf{h}}} + \mathbf{I}_N \right) \right] + N n_R \log \left(\pi e \sigma_n^2 \right) \qquad (8.50)$$

where $\mathbf{Z} = \mathrm{diag}(z(1), \ldots, z(N))$ and

$$z(k) = \sum_{j=1}^{n_T} \frac{p_j}{\sigma_x^2} \phi_j |x_j(k)|^2. \qquad (8.51)$$

For (a), we have used the following inequality that has been proven in [102, Lemma 4]

$$\det(\mathbf{I}_N + \mathbf{A} \odot \mathbf{B}) \geq \det \left(\mathbf{I}_N + (\mathbf{I}_N \odot \mathbf{A})\mathbf{B} \right) \qquad (8.52)$$

for $\mathbf{A}$ and $\mathbf{B}$ being $N \times N$ nonnegative definite Hermitian matrices. Here $\odot$ denotes the Hadamard product.

Now, following exactly the same approach as in the SISO case based on the asymptotic equivalence of Toeplitz and circulant matrices, see Section 3.3.2, we substitute the Toeplitz matrix $\mathbf{R}_{\bar{\mathbf{h}}}$ by a circulant matrix $\mathbf{C}_{\bar{\mathbf{h}}}$ with the eigenvalue distribution given in (3.44) which corresponds to a rectangular PSD of the channel fading process for $N \to \infty$. As in Section 3.3.2, the spectral decomposition of $\mathbf{C}_{\bar{\mathbf{h}}}$ is given by

$$\mathbf{C}_{\bar{\mathbf{h}}} = \mathbf{F} \tilde{\boldsymbol{\Lambda}}_h \mathbf{F}^H \qquad (8.53)$$

with $\mathbf{F}$ being unitary and $\tilde{\boldsymbol{\Lambda}}_h$ being diagonal and containing the eigenvalues given in (3.44).

Based on this, the entropy rate $h'(\mathbf{y}|\mathbf{x})$ in (8.50) can be lower-bounded by

$$h'(\mathbf{y}|\mathbf{x}) \geq \lim_{N \to \infty} \frac{1}{N} \left[\mathrm{E_x} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \mathbf{Z} \mathbf{R}_{\bar{\mathbf{h}}} + \mathbf{I}_N \right) \right] + N n_R \log \left(\pi e \sigma_n^2 \right) \right]$$

$$\overset{(a)}{=} \lim_{N \to \infty} \frac{1}{N} \left[\mathrm{E_x} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \mathbf{Z} \mathbf{C}_{\bar{\mathbf{h}}} + \mathbf{I}_N \right) \right] + N n_R \log \left(\pi e \sigma_n^2 \right) \right]$$

$$\overset{(b)}{=} \lim_{N \to \infty} \frac{1}{N} \mathrm{E_x} \left[\sum_{i=1}^{n_R} \log \det \left(\frac{\psi_i}{\sigma_n^2} \tilde{\boldsymbol{\Lambda}}_h \mathbf{F}^H \mathbf{Z} \mathbf{F} + \mathbf{I}_N \right) \right] + n_R \log \left(\pi e \sigma_n^2 \right)$$

$$\geq \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{n_R} \sum_{k=1}^{\lfloor 2 f_d N \rfloor} \mathrm{E_x} \log \left(\frac{\psi_i}{\sigma_n^2} \frac{\sigma_h^2 z(k)}{2 f_d} + 1 \right) + n_R \log(\pi e \sigma_n^2)$$

$$= 2 f_d \sum_{i=1}^{n_R} \mathrm{E_x} \log \left(\frac{\psi_i}{\sigma_n^2} \frac{\sigma_h^2}{2 f_d} \sum_{j=1}^{n_T} \frac{p_j}{\sigma_x^2} \phi_j |x_j(1)|^2 + 1 \right) + n_R \log(\pi e \sigma_n^2)$$

$$= h'_L(\mathbf{y}|\mathbf{x}) \qquad (8.54)$$

where for (a) we have used the asymptotic equivalence of $\frac{\psi_i}{\sigma_n^2}\mathbf{Z}\mathbf{R}_{\bar{\mathbf{h}}} + \mathbf{I}_N$ and $\frac{\psi_i}{\sigma_n^2}\mathbf{Z}\mathbf{C}_{\bar{\mathbf{h}}} + \mathbf{I}_N$, which can be shown analogously as in the SISO case in Section 3.3.2. In addition, for (b) we have used the spectral decomposition of $\mathbf{C}_{\bar{\mathbf{h}}}$ and (3.25). The remaining steps are analogous to the ones in the SISO case, see Lemma 3.1 and cf. (3.59)–(3.60). In the second last line of (8.54) for $x_j(k)$ the time instant $k = 1$ is chosen as an arbitrary time instant.

With (8.54) we have found a general expression for the lower bound on $h'(\mathbf{y}|\mathbf{x})$ for i.d. input distributions. As the derivation of this bound is not based on a peak power constraint, it enables us to give upper bounds on the achievable rate for proper Gaussian input symbols.

8.2.3 The Achievable Rate

Based on the derived lower and upper bounds on $h'(\mathbf{y})$ and $h'(\mathbf{y}|\mathbf{x})$ we now give upper and lower bounds on the achievable rate with temporally i.i.d. zero-mean proper Gaussian inputs and the constraint on a fixed spatial input signal covariance matrix as discussed above. Furthermore, we also give bounds on the achievable rate with temporally i.d. peak power constrained input symbols.

8.2.3.1 Upper Bound

Gaussian Input Distribution

Obviously, with (3.10) an upper bound on the achievable rate is given by

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \tag{8.55}$$

with $h'_U(\mathbf{y})$ and $h'_L(\mathbf{y}|\mathbf{x})$ given by (8.42) and (8.54). Now, we give an upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ for the special case of a zero-mean proper Gaussian input distribution, with i.i.d. input elements in temporal and also in the eigenmode domain given by the inputs of the spatially decorrelated system model in (8.29). This means that the elements of $\check{\mathbf{x}}(k)$ are i.i.d.. Due to the transformation in (8.29) this corresponds to the capacity-achieving input distribution in case of perfect channel state information at the receiver side. In this case and for an average power of $\alpha\sigma_x^2$ per element of $\check{\mathbf{x}}(k)$, i.e., per eigenmode the lower bound on $h'(\mathbf{y}|\mathbf{x})$ in (8.54) becomes

$$h'_L(\mathbf{y}|\mathbf{x})\big|_{\mathrm{PG}} = 2f_d \sum_{i=1}^{n_R} \int_0^\infty \log\left(\frac{\psi_i}{\sigma_n^2}\frac{\sigma_h^2}{2f_d}\alpha \sum_{j=1}^{n_T} p_j\phi_j u_j + 1\right) \prod_{j=1}^{n_T} e^{-u_j} du_j$$
$$+ n_R \log(\pi e \sigma_n^2) \tag{8.56}$$

where the index PG indicates i.i.d. zero-mean proper Gaussian input symbols. With (8.56) we get

$$\sup_{\alpha\in[0,1]} \mathcal{I}'(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} \leq \sup_{\alpha\in[0,1]} \left\{ \sum_{i=1}^{n_R} \log\left(\psi_i\alpha\rho\frac{\sum_{j=1}^{n_T}(p_j\phi_j)}{P}+1\right) \right.$$
$$\left. - 2f_d \sum_{i=1}^{n_R} \int_0^\infty \log\left(\frac{\psi_i}{2f_d}\alpha\rho \sum_{j=1}^{n_T} \frac{p_j}{P}\phi_j u_j + 1\right) \prod_{j=1}^{n_T} e^{-u_j} du_j \right\}$$
$$\overset{(a)}{=} \sum_{i=1}^{n_R} \log\left(\psi_i\rho\frac{\sum_{j=1}^{n_T}(p_j\phi_j)}{P}+1\right)$$
$$- 2f_d \sum_{i=1}^{n_R} \int_0^\infty \log\left(\frac{\psi_i}{2f_d}\rho \sum_{j=1}^{n_T} \frac{p_j}{P}\phi_j u_j + 1\right) \prod_{j=1}^{n_T} e^{-u_j} du_j$$
$$= \mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} \tag{8.57}$$

where (a) corresponds to the observation that the upper bound is maximized by using the maximum average power P, which can be shown analogous to the SISO case discussed in Section 3.4.1.2. To the best of our knowledge the upper bound in (8.57) is new.

Corresponding to (3.87) in the SISO case, the upper bound in (8.57) can also be modified based on the coherent capacity which forms an upper bound to the noncoherent capacity. The coherent capacity in the MIMO scenario is given by (8.64) and for the special case of a spatially uncorrelated channel by (8.66), see below.

Peak Power Constraint

For the case of a peak power constraint, we have to calculate the following supremum

$$\sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,\ i.d.}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,\ i.d.}}} \left\{ h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \right\}. \tag{8.58}$$

Following in principle the same bounding approach as for the SISO case in Section 3.4.1.1, the supremum in (8.58) yields

$$\sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO, \, i.d.}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \le \mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{P_{\mathrm{peak}}}$$

$$= \sup_{\alpha\in[0,1]} \left\{ \sum_{i=1}^{n_R} \log\left(\alpha\psi_i\rho\frac{\sum_{j=1}^{n_T} p_j\phi_j}{P} + 1 \right) \right.$$
$$\left. - 2f_d\frac{\alpha}{\beta}\frac{\sum_{j=1}^{n_T} p_j\phi_j}{n_T \max_j\{p_j\phi_j\}} \sum_{i=1}^{n_R} \log\left(\frac{\psi_i}{2f_d}\rho\beta n_T\frac{\max_j\{p_j\phi_j\}}{P} + 1 \right) \right\}$$

$$(8.59)$$

with the nominal peak-to-average power ratio β defined by, cf. (3.75)

$$\beta = \frac{P_{\mathrm{peak}}}{P}. \qquad (8.60)$$

For the derivation of (8.59) we have used that

$$\inf_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,i.d.}}|\alpha} \mathrm{E_x}\log\left(\frac{\psi_i}{\sigma_n^2}\frac{\sigma_h^2}{2f_d}\sum_{j=1}^{n_T}\frac{p_j}{\sigma_x^2}\phi_j|x_j(1)|^2 + 1 \right)$$

$$= \inf_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,i.d.}}|\alpha} \int_{\mathbf{\check{x}}} \frac{\log\left(\frac{\psi_i}{\sigma_n^2}\frac{\sigma_h^2}{2f_d}\sum_{j=1}^{n_T}\frac{p_j}{\sigma_x^2}\phi_j|x_j(1)|^2 + 1 \right)}{\sum_{j=1}^{n_T}\frac{p_j}{\sigma_x^2}\phi_j|x_j(1)|^2} \sum_{j=1}^{n_T}\frac{p_j}{\sigma_x^2}\phi_j|x_j(1)|^2 p(\mathbf{\check{x}})d\mathbf{\check{x}}$$

$$\overset{(a)}{=} \frac{\log\left(\frac{\psi_i}{\sigma_n^2}\frac{\sigma_h^2}{2f_d}\frac{\max_j\{p_j\phi_j\}}{\sigma_x^2}P_{\mathrm{peak}} + 1 \right)}{\frac{\max_j\{p_j\phi_j\}}{\sigma_x^2}P_{\mathrm{peak}}}\alpha\sum_{j=1}^{n_T}p_j\phi_j$$

$$= \frac{\alpha}{\beta}\frac{1}{n_T}\frac{\sum_{j=1}^{n_T}p_j\phi_j}{\max_j\{p_j\phi_j\}}\log\left(\frac{\psi_i}{2f_d}\rho\beta\frac{\max_j\{p_j\phi_j\}}{\sigma_x^2} + 1 \right) \qquad (8.61)$$

where (a) results from a similar argumentation as in (3.71). Furthermore, the set $\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,i.d.}}|\alpha$ corresponds to $\mathcal{P}^{\mathrm{peak}}_{\mathrm{MIMO,i.d.}}$ in (8.37) but with an average transmit power fixed to αP with equality.

Notice that the calculation of the supremum in (8.58) is based on the assumption of a fixed spatial input covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$ in (8.11) and, therefore, on a fixed $\tilde{\mathbf{P}}$ in the transformed channel in (8.29). Hence, the powers p_j in (8.59) are fixed. Furthermore, we assume that they are chosen such that they would be optimal in case of a coherent channel. For a spatially uncorrelated channel (8.59) simplifies to

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{P_{\mathrm{peak}}} = \sup_{\alpha\in[0,1]} \left\{ n_R\log\left(\alpha\rho + 1 \right) - 2f_d\frac{\alpha}{\beta}n_R\log\left(\frac{1}{2f_d}\rho\beta + 1 \right) \right\} \qquad (8.62)$$

as in this case $\psi_i = 1$, $\phi_i = 1$, and $p_j = P/n_T$.

With (8.59), we have found an upper bound on the achievable rate with temporally i.d. distributed input symbols with a peak power constraint. Here it has to be stressed again that the transmit covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$, see (8.11) is not optimized, but taken as it would be optimal in case of perfect CSI at the receiver. Notice that this upper bound only holds for a rectangular PSD of the channel fading process. As in the case of Gaussian input distributions, the upper bound in (8.59) can be modified using the coherent upper bound, cf. (3.87).

8.2.3.2 Lower Bound

No Peak Power Constraint

Now, we give a lower bound on the achievable rate with non peak power constrained input symbols. Obviously, as a lower bound on the achievable rate, this bound is also a lower bound on the capacity. Furthermore, this lower bound holds also when restricting to i.i.d. zero-mean proper Gaussian input symbols.

Using (8.38) and (8.48), a lower bound on the capacity that can be achieved with temporally i.d. input symbols is given by

$$
\sup_{\mathcal{P}_{\text{MIMO, i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \geq \sup_{\mathcal{P}_{\text{MIMO, i.d.}}} \{h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x})\}
$$

$$
= \sup_{\mathcal{P}_{\text{MIMO, i.d.}}} \{\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) - h'_U(\mathbf{y}|\mathbf{x})\}
$$

$$
= \sup_{\mathcal{P}_{\text{MIMO, i.d.}}} \left\{ \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i \alpha p_j \phi_j}{\sigma_n^2} S_h(f) + 1 \right) df \right\}
$$

$$
= \sup_{\alpha \in [0,1]} \left\{ \sup_{\mathcal{P}_{\text{MIMO, i.d.}}|\alpha} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i \alpha p_j \phi_j}{\sigma_n^2} S_h(f) + 1 \right) df \right\}
$$

$$
\tag{8.63}
$$

where $\mathcal{P}_{\text{MIMO, i.d.}}|\alpha$ corresponds to $\mathcal{P}_{\text{MIMO, i.d.}}$, but with a maximum average transmit power of αP.

The first term on the RHS of (8.63), i.e., $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is the mutual information rate in case of perfect channel knowledge. It is well known that the supremum of this term, i.e., the capacity, is achieved for i.i.d. zero-mean proper Gaussian input vectors over time and is given by

$$\sup_{\mathcal{P}_{\text{MIMO, i.d.}|\alpha}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) = \sup_{\mathcal{P}_{\text{MIMO}|\alpha}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$$

$$= \sup_{\{\mathbf{R}_{\tilde{\mathbf{x}}}|\text{E}[(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)]\leq\alpha P\}} \text{E}_{\tilde{\mathbf{H}}} \log\det\left(\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{H}}^H\right) \tag{8.64}$$

where $\mathcal{P}_{\text{MIMO}|\alpha}$ corresponds to $\mathcal{P}_{\text{MIMO}}$ but with a maximum average power of αP. Notice, we express $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ based on the genuine system model in (8.1), i.e., not based on the spatially decorrelated model in (8.29).

Thus, we get

$$\sup_{\mathcal{P}_{\text{MIMO, i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x})$$

$$\geq \sup_{\alpha[0,1]} \left\{ \sup_{\{\mathbf{R}_{\tilde{\mathbf{x}}}|\text{E}[(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)]\leq\alpha P\}} \text{E}_{\tilde{\mathbf{H}}} \log\det\left(\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{H}}^H\right) \right.$$

$$\left. - \sum_{i=1}^{n_R}\sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i\alpha p_j\phi_j}{\sigma_n^2}S_h(f) + 1\right) df \right\}$$

$$\overset{(a)}{\geq} \sup_{\{\mathbf{R}_{\tilde{\mathbf{x}}}|\text{E}[(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)]\leq P\}} \text{E}_{\tilde{\mathbf{H}}} \log\det\left(\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{H}}^H\right)$$

$$- \sum_{i=1}^{n_R}\sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i p_j\phi_j}{\sigma_n^2}S_h(f) + 1\right) df$$

$$= \mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\text{PG}} \tag{8.65}$$

where for (a) we choose $\alpha = 1$, i.e., use the maximum average transmit power, which does not maximize the argument of the supremum over α in general. In addition, the last equality holds as the given lower bound is achievable with i.i.d. (in temporal domain) zero-mean proper Gaussian input symbol vectors $\tilde{\mathbf{x}}(k)$.

Expressions for the coherent capacity, i.e., the first term on the RHS of (8.65) will be recalled in the following. Thus, with (8.65) we have found a lower bound on the capacity. Notice, the modification of the lower bound (8.65) based on the nonnegativity of mutual information that has been used in (3.94) for the SISO case can also be applied in the MIMO case.

Spatially Uncorrelated MIMO Channel

In [120] it has been shown that in case of a coherent and spatially uncorrelated MIMO fading channel spatially independent transmit symbols, i.e., $\mathbf{R}_{\tilde{\mathbf{x}}} = \sigma_x^2\mathbf{I}_{n_T}$ are capacity-achieving. The capacity is then given by the following closed form solution of (8.64) for $\alpha = 1$

$$\sup_{\mathcal{P}_{\text{MIMO}}} \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) = \int_0^\infty \log\left(1 + \frac{\rho}{n_T}\lambda\right) \sum_{k=0}^{m-1} \frac{k![L_k^{n-m}(\lambda)]^2}{(k+n-m)!} \lambda^{n-m} e^{-\lambda} d\lambda$$

(8.66)

where $m = \min\{n_T, n_R\}$, $n = \max\{n_T, n_R\}$, L_j^i are the associated Laguerre polynomials, and ρ is the SNR as defined in (8.13).

Introducing (8.66) into (8.65) results in a lower bound on the capacity with non-peak power constrained input symbols in case of a spatially uncorrelated channel.

Spatially Correlated MIMO Channel

For the case of spatially correlated coherent fading channels, the capacity-achieving input covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$ has already been discussed in Section 8.1.2.1.

The capacity for this case has been studied in [123] based on an asymptotic analysis. This analysis is asymptotic in the number of transmit and receive antennas and is based on means of random matrix theory. The coherent capacity converges almost surely to

$$\sup_{\mathcal{P}_{\text{MIMO}}} \frac{1}{n_R}\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$$
$$= \mu\mathrm{E}\left[\log(1 + \rho\Lambda\Gamma(\rho))\right] + \mathrm{E}\left[\log(1 + \rho\Lambda_R\Upsilon(\rho))\right] - \mu\rho\Gamma(\rho)\Upsilon(\rho) \quad (8.67)$$

where $\mu = \frac{n_T}{n_R}$ and

$$\Gamma(\rho) = \frac{1}{\mu}\mathrm{E}\left[\frac{\Lambda_R}{1 + \rho\Lambda_R\Upsilon(\rho)}\right] \tag{8.68}$$

$$\Upsilon(\rho) = \mathrm{E}\left[\frac{\Lambda}{1 + \rho\Lambda\Gamma(\rho)}\right] \tag{8.69}$$

with the expectation over Λ and Λ_R whose distributions are given by the asymptotic empirical eigenvalue distribution of $\tilde{\boldsymbol{\Phi}}\mathbf{P}$ and $\tilde{\boldsymbol{\Psi}}$. In [123] it has been shown that this asymptotic approximation already becomes tight for a relatively small number of antennas. Therefore, (8.67) only holds explicitly for an asymptotic number of antennas, and approximately already for a few number of transmit and receive antennas [123].

Peak Power Constraint

Obviously, the lower bound in (8.65) does not hold when we introduce a peak power constraint. Corresponding to the SISO case discussed in Section 3.4.2.2,

we can also give the following lower bound on the achievable rate with a peak power constraint

$$
\begin{aligned}
\sup_{\mathcal{P}_{\text{MIMO, i.d.}}^{\text{peak}}} \mathcal{I}_L'(\mathbf{y}; \mathbf{x}) &= \sup_{\mathcal{P}_{\text{MIMO, i.d.}}^{\text{peak}}} \{h_L'(\mathbf{y}) - h_U'(\mathbf{y}|\mathbf{x})\} \\
&= \sup_{\mathcal{P}_{\text{MIMO, i.d.}}^{\text{peak}}} \{\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x}, \mathbf{h}) - h_U'(\mathbf{y}|\mathbf{x})\} \\
&\geq \mathcal{I}\left(\check{\mathbf{y}}; \check{\mathbf{x}}|\tilde{\mathbf{h}}\right)\Big|_{\text{CM},P} - \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\psi_i p_j \phi_j}{\sigma_n^2} S_h(f) + 1\right) df \\
&= \mathcal{I}_L'(\mathbf{y}; \mathbf{x})\Big|_{P_{\text{peak}}} \tag{8.70}
\end{aligned}
$$

where $\mathcal{I}\left(\check{\mathbf{y}}; \check{\mathbf{x}}|\tilde{\mathbf{h}}\right)\big|_{\text{CM},P}$ is the coherent mutual information for input symbol vectors $\check{\mathbf{x}}(k)$ consisting of i.i.d. zero-mean circularly symmetric constant modulus input symbols with power $\sigma_x^2 = P/n_T$, cf. (8.29). Thus, with $\mathcal{I}_L'(\mathbf{y}; \mathbf{x})\big|_{P_{\text{peak}}}$ we have found a lower bound on the peak power constrained capacity that is achievable with temporally i.d. input symbols. However, as the bound in (8.70) corresponds to a lower bound for constant modulus signaling it is not tight in general. Like in the SISO case, there is no closed form solution available for the coherent capacity with constant modulus inputs, i.e., the first term in (8.70) has to be calculated numerically, see, e.g., [49] for PSK signaling.

For the case of frequency-selective MIMO fading channels, a lower bound that is related to (8.70) is known from [102], see Section 8.2.3.6 for a detailed discussion.

8.2.3.3 Numerical Evaluation for Gaussian Inputs

Fig. 8.2a to Fig. 8.2c show the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs and no spatial channel correlation for an SNR of 6 dB for an equal number of transmit and receive antennas, one receive antenna and various transmit antennas, and one transmit antenna and various receive antennas. Obviously, these bounds become very loose for an increasing number of transmit and receive antennas.

Fig. 8.2d shows exemplarily for $n_T = n_R = 5$ the effect of the SNR on the bounds on the achievable rate, showing that the gap between the upper and lower bound decreases with an decreasing SNR.

For $f_d \to 0$ the well known effects of the number of transmit and receive antennas in case of a coherent channel, e.g., described in [120], can be observed. That means that for an equal number of transmit and receive antennas the capacity approximately increases linearly with the number of antennas. For one receive antenna the capacity slightly increases with an increasing number of transmit antennas, approaching the AWGN channel capacity for $n_T \to \infty$.

For one transmit antenna and with an increasing number of receive antennas, the capacity increases approximately logarithmically. These results are obvious, as in case of $f_d \to 0$ the second term in the lower bound on the capacity in (8.65) approaches zero and thus (8.65) approaches the coherent capacity. Obviously, the lower bound becomes tight for $f_d \to 0$.

For larger f_d it is in general difficult to give statements on the achievable rate as the bounds are not tight.

In general, it has to be considered that all bounds shown in Fig. 8.2 hold only while using the input distribution being capacity-achieving in case of a coherent channel. In the case of a spatially uncorrelated channel as considered in Fig. 8.2 this means that we assume i.i.d. input symbols in time and over the transmit antennas, which are all zero-mean proper Gaussian. In general, these input distributions will not be capacity-achieving in the noncoherent scenario. And if we deviate from this assumption, we could also enhance the bounds on the achievable rate. Therefore, consider, e.g., Fig. 8.2a. Obviously, the lower bound for $n_T = 1$ is larger than the lower bound for $n_T = 5$ for $f_d > 0.05$. This means, that when switching of transmit antennas, we get a larger lower bound on the achievable rate. But switching of antennas means to modify the input distribution, which is not captured by our bounds as we initially made the assumption that we use the input distribution that is capacity-achieving in case of a coherent channel. Likewise, studying Fig. 8.2b, it can be observed that the upper bound on the achievable rate with $n_T = 5$ transmit antennas is smaller than the upper bound for $n_T = 1$ transmit antennas for $f_d > 0.03$. However, as the upper bounds and the lower bounds on the achievable rate are not tight, the given bounds are not sufficient to give statements on the optimum number of transmit antennas depending on the channel dynamics.

8.2.3.4 Tightness

In this section, we discuss the tightness of the upper and the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs for the special case of a spatially uncorrelated channel.

In contrast to the SISO case discussed in Section 3.4.3.2, it is difficult to evaluate the tightness of the upper and the lower bound on the achievable rate given in (8.57) and (8.65) based on the tightness of the upper and lower bounds on $h'(\mathbf{y})$ and $h'(\mathbf{y}|\mathbf{x})$. Therefore, we evaluate the tightness based on the following separation of the mutual information rate, given by inserting (3.8) into (3.7)

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) - \mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x}) + \mathcal{I}'(\mathbf{y};\mathbf{h}). \qquad (8.71)$$

The first term on the RHS of (8.71) $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$, which is the coherent mutual information rate, is known and given in (8.66).
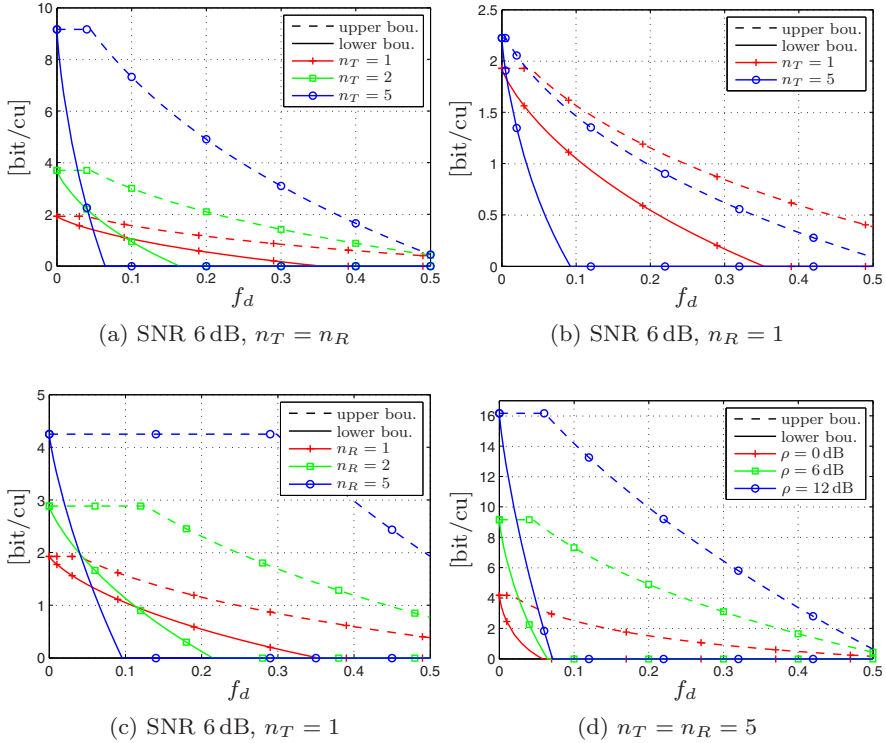
Fig. 8.2: **Achievable rates with i.i.d. zero-mean proper Gaussian input distribution**, spatially uncorrelated MIMO channel; upper bound (8.57) in combination with coherent capacity, cf. (3.87)/(8.66); lower bound (8.65) in combination with nonnegativity of mutual information, cf. (3.94); rectangular PSD $S_h(f)$

On the one hand, we can express the lower bound on the achievable rate in (8.63) as follows

$$
\begin{aligned}
\mathcal{I}'_L(\mathbf{y}; \mathbf{x}) &= h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x}) \\
&= \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) - \mathcal{I}'_U(\mathbf{y}; \mathbf{h}|\mathbf{x}) - h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) \\
&= \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \mathcal{I}'_U(\mathbf{y}; \mathbf{h}|\mathbf{x})
\end{aligned}
\tag{8.72}
$$

where $\mathcal{I}'_U(\mathbf{y}; \mathbf{h}|\mathbf{x})$ is an upper bound on $\mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x})$. As the comparison with (8.71) shows, for the lower bound on the achievable rate $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$, we use an upper bound on $\mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x})$ and the following lower bound

$$
\mathcal{I}'_L(\mathbf{y}; \mathbf{h}) = 0.
\tag{8.73}
$$

On the other hand, we can express the upper bound in (8.57) by

$$
\begin{aligned}
\mathcal{I}'_U(\mathbf{y};\mathbf{x}) &= h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \\
&= h'_U(\mathbf{y}) - h'(\mathbf{y}|\mathbf{h}) - h'_L(\mathbf{y}|\mathbf{x}) + h'(\mathbf{y}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) - h'(\mathbf{y}|\mathbf{x},\mathbf{h}) \\
&= \mathcal{I}'_U(\mathbf{y};\mathbf{h}) - \mathcal{I}'_L(\mathbf{y};\mathbf{h}|\mathbf{x}) + \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})
\end{aligned}
\tag{8.74}
$$

where $\mathcal{I}'_L(\mathbf{y};\mathbf{h}|\mathbf{x})$ is a lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ and $\mathcal{I}'_U(\mathbf{y};\mathbf{h})$ is an upper bound on $\mathcal{I}'(\mathbf{y};\mathbf{h})$. The upper bound on $\mathcal{I}'(\mathbf{y};\mathbf{h})$ is given by

$$
\begin{aligned}
\mathcal{I}'_U(\mathbf{y};\mathbf{h}) &= h'_U(\mathbf{y}) - h'(\mathbf{y}|\mathbf{h}) \\
&= h'_U(\mathbf{y}) - \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) - h'(\mathbf{y}|\mathbf{x},\mathbf{h}).
\end{aligned}
\tag{8.75}
$$

For $h'_U(\mathbf{y})$, we use the upper bound in (8.43) given by the fact that the entropy rate is maximized for proper Gaussian random variables.

The upper and lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ in (8.72) and (8.74) can be expressed as

$$
\mathcal{I}'_L(\mathbf{y};\mathbf{h}|\mathbf{x}) = h'_L(\mathbf{y}|\mathbf{x}) - h'(\mathbf{y}|\mathbf{h},\mathbf{x})
\tag{8.76}
$$

$$
\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x}) = h'_U(\mathbf{y}|\mathbf{x}) - h'(\mathbf{y}|\mathbf{h},\mathbf{x})
\tag{8.77}
$$

where $h'_U(\mathbf{y}|\mathbf{x})$ and $h'_L(\mathbf{y}|\mathbf{x})$ have been calculated in (8.48) and (8.56).

Now consider that the upper bound $\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x})$ is used to lower-bound $\mathcal{I}'(\mathbf{y};\mathbf{x})$ in (8.72). As mutual information is always nonnegative, a simple lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{x})$ is given by zero. Therefore, for the final lower bound $\mathcal{I}'_{L_{\mathrm{mod}}}(\mathbf{y};\mathbf{x})$, cf. (3.94), we use the upper bound $\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x})$ only if it is smaller than $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$, see (8.72). Therefore, we can introduce the following auxiliary quantity

$$
\mathcal{I}'_{U_{\mathrm{useful}}}(\mathbf{y};\mathbf{h}|\mathbf{x}) = \min\left\{\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x}), \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})\right\}.
\tag{8.78}
$$

Note that $\mathcal{I}'_{U_{\mathrm{useful}}}(\mathbf{y};\mathbf{h}|\mathbf{x})$ is not an upper bound on $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$.

Fig. 8.3 shows the upper and lower bounds on $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ and an approximative (finite observation interval length) numerical evaluation of $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ for a spatially uncorrelated channel, i.i.d. zero-mean proper Gaussian inputs, $n_T = n_R = 5$ transmit and receive antennas, and an SNR of 6 dB. This figure clearly shows that the upper bound $\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x})$ is very close to $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ for the channel dynamics f_d where $\mathcal{I}'_{U_{\mathrm{useful}}}(\mathbf{y};\mathbf{h}|\mathbf{x})$ is not anyway given by $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$, see (8.78). Therefore, the gap between the lower bound $\mathcal{I}'_L(\mathbf{y};\mathbf{x})$ and the actual achievable rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ is mainly given by the term $\mathcal{I}'(\mathbf{y};\mathbf{h})$ which we lower-bound by zero, see (8.73). Therefore, to get a tighter lower bound on the achievable rate we need a tighter lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{h})$ than the one given in (8.73). That means that a tighter lower bound on the output entropy rate $h'(\mathbf{y})$ would be required.

To evaluate the tightness of the upper bound $\mathcal{I}'_U(\mathbf{y};\mathbf{x})$ we have to study how tight $\mathcal{I}'_U(\mathbf{y};\mathbf{h})$ and $\mathcal{I}'_L(\mathbf{y};\mathbf{h}|\mathbf{x})$ are, see (8.74). Fig. 8.3 shows $\mathcal{I}'_L(\mathbf{y};\mathbf{h}|\mathbf{x})$ in
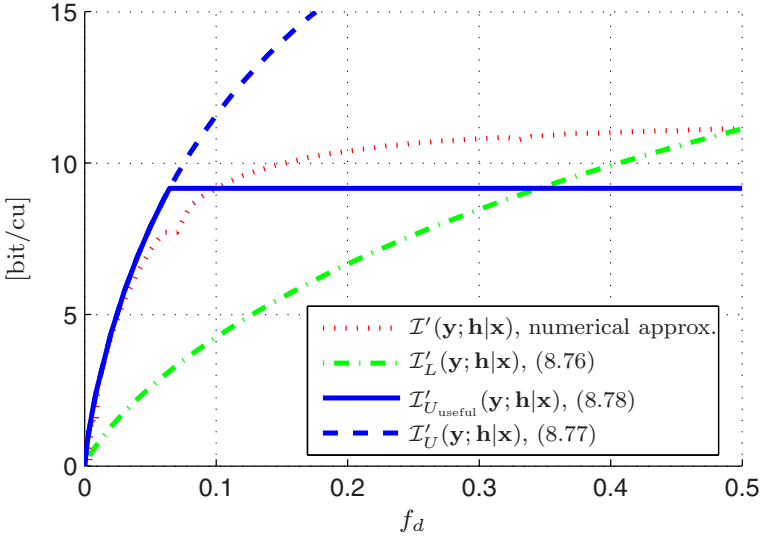
Fig. 8.3: Tightness of bounds on $\mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x})$, SNR 6 dB, $n_T = n_R = 5$; notice, $\mathcal{I}'_{U_{\text{useful}}}(\mathbf{y}; \mathbf{h}|\mathbf{x})$ is not an upper bound on $\mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x})$; rectangular PSD $S_h(f)$

comparison to its approximative numerical evaluation. Obviously, for $f_d = 0$ and $f_d = 0.5$ $\mathcal{I}'_L(\mathbf{y}; \mathbf{h}|\mathbf{x})$ becomes tight, while in between there is a gap. For the upper bound $\mathcal{I}'_U(\mathbf{y}; \mathbf{h})$ in (8.75) we use the upper bound $h'_U(\mathbf{y})$ in (8.43) and thus it is based on the entropy maximizing characteristic of proper Gaussian random variables. We know that this bound is increasingly loose with an increasing number of receive antennas and depending on the channel dynamics. For small channel dynamics it is very loose due to the fact that its calculation $h'_U(\mathbf{y})$ assumes that the receive symbols at different time instances are independent. This assumption has been taken implicitly, as the upper bound is based on proper Gaussian random variables. In fact, the receive symbols at the different time instances are uncorrelated. As proper Gaussian random variables that are uncorrelated are also independent, during upper bounding we discard all dependencies between the received symbols at different time instances, while in fact the observations $\tilde{\mathbf{y}}(k)$ at different time instances are not independent. But also for the extreme case of an uncorrelated channel, i.e., $f_d = 0.5$, the upper bound $h'_U(\mathbf{y})$ is still loose, as due to the bounding based on proper Gaussian random variables the observations of different receive antennas are assumed to be independent. This is not the case in reality as the observations at the different receive antennas depend on common transmit symbols. In fact, the observations of different receive antennas at the same time instant are uncorrelated, but not independent.

In conclusion it can be stated that for tighter upper and lower bounds on the achievable rate tighter upper and lower bounds on the channel output entropy rate $h'(\mathbf{y})$ are strongly required. Furthermore, a tighter lower bound on $\mathcal{I}'(\mathbf{y};\mathbf{h}|\mathbf{x})$ would be valuable, whereas the tightness of $\mathcal{I}'_U(\mathbf{y};\mathbf{h}|\mathbf{x})$ is not a major issue. For the SISO case we already tried to give an enhanced upper bound on $h'(\mathbf{y})$, see Section 4.2.2. However, already in the SISO case we did not get to a solution for an enhanced upper bound on $h'(\mathbf{y})$.

8.2.3.5 Effect of Antenna Correlation

In this section, we show the effect of spatial antenna correlation on the achievable rate. For zero-mean proper Gaussian input symbols in Fig. 8.4 the bounds on the achievable rate for a MIMO channel with $n_T = n_R = 5$ transmit and receive antennas have been plotted for different SNRs and different spatial antenna correlations including a comparison to the case of uncorrelated transmit and receive antennas. For the spatially correlated channel the asymptotic approximation in (8.67) has been used. Furthermore, the spatial transmit signal covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$ has been chosen as discussed in Section 8.1.2.1. The antenna correlation matrices $\mathbf{\Theta}_T$ and $\mathbf{\Theta_R}$ are given by

$$\left[\mathbf{\Theta}_{T/R}\right]_{k,l} = e^{-0.05d_{T/R}^2(k-l)^2} \tag{8.79}$$

where d_T is the antenna spacing between the individual transmit antennas of a linear antenna array, and analogous d_R is the antenna spacing of the linear receive antenna array both in wavelengths. This model has also been used in [122], however, only for spatial correlation at the transmitter side.

Fig. 8.4a shows the effect of spatial antenna correlation for an SNR of 6 dB. Both transmit and receive antennas are correlated with $d_T = d_R = 1$. Obviously, for small channel dynamics ($f_d < 0.01$) it can be stated that the antenna correlation reduces the achievable rate. Fig. 8.4b and Fig. 8.4c show the cases where spatial correlation exists only at the transmitter or the receiver side. Also in these two cases the antenna correlation reduces the achievable rate for small channel dynamics. Due to the fact that the bounds are rather loose, it is not possible to give statements on the effect of spatial channel correlation in general. As the upper and lower bound on the mutual information get closer for smaller SNR, in Fig. 8.4d to Fig. 8.4f we plot the upper and lower bounds on the achievable rate for a very low SNR of -6 dB and for different spatial antenna correlations. Fig. 8.4d shows the achievable rate for spatial correlation at the transmit and the receive antennas, Fig. 8.4e shows the achievable rate for the case of spatially correlated receive antennas and uncorrelated transmit antennas, and Fig. 8.4f shows the achievable rate for the case of spatially uncorrelated receive antennas and spatially correlated transmit antennas. First we look at very small channel dynamics. For a very small SNR of -6 dB it can be observed that spatial correlation at the

receiver side reduces the achievable rate, see Fig. 8.4e, spatial correlation at
the transmitter side significantly increases the achievable rate, see Fig. 8.4f,
and Fig. 8.4d shows a superposition of both effects. General statements on
the effect of channel correlation for higher channel dynamics and higher SNR
are rather difficult.

For the case of perfect channel state information at the receiver side and
covariance information at the transmitter side, in [57] the following state-
ments are given

- spatial correlation at the receiver side always decreases capacity
- spatial correlation at the transmitter side decreases capacity for high SNR
  and is beneficial for low SNR.

In contrast to [57], we consider the case where the receiver does not know
the realization of the channel, but only its statistics. The assumption on
the knowledge of the channel at the transmitter side are the same in [57]
and the present work, i.e., assuming knowledge of the channel statistics. For
asymptotically small channel dynamics, the results we observe correspond to
the behavior shown in [57] for the case of perfect CSI at the receiver, which
clearly must be the case, as for asymptotically small channel dynamics, the
channel can be estimated arbitrarily well.

To conclude, it can be said that due to the gap between upper and lower
bounds on the achievable rate, it is not possible to get a general understanding
on the effect of spatial correlation. E.g., for the wideband case, it has been
stated in [102] that for sufficiently large bandwidths transmit and receive
correlation is beneficial, which is rather different to the coherent setting,
where receive antenna correlation is always detrimental. This behavior is
explained in [102] by the effect that channel uncertainty decreases due to
receive antenna correlation. Due to the lack of tightness of our bounds, we do
not observe this behavior here for the flat-fading case, however, e.g., Fig. 8.4e
is not in contradiction to this.

8.2.3.6 Relation to Bounds in [102]

A comparison of our upper and lower bounds on the achievable rate shows
some similarity to the upper and lower bounds on the peak power constrained
capacity given in [102] for frequency-selective wideband channels. Besides the
difference of flat and frequency selectivity, the major difference lies in the up-
per bounding technique. The derivation of the upper bound given in [102] is
based on the peak power constraint. Therefore, the bounding technique that
has been used in [102] for the upper bound cannot be applied when consider-
ing non peak power constrained proper Gaussian inputs. For the derivation of
the upper bound on the achievable rate, we have used a completely different
bounding technique, which does not rely on the peak power constraint. This
enables to upper bound the achievable rate with zero-mean proper Gaus-
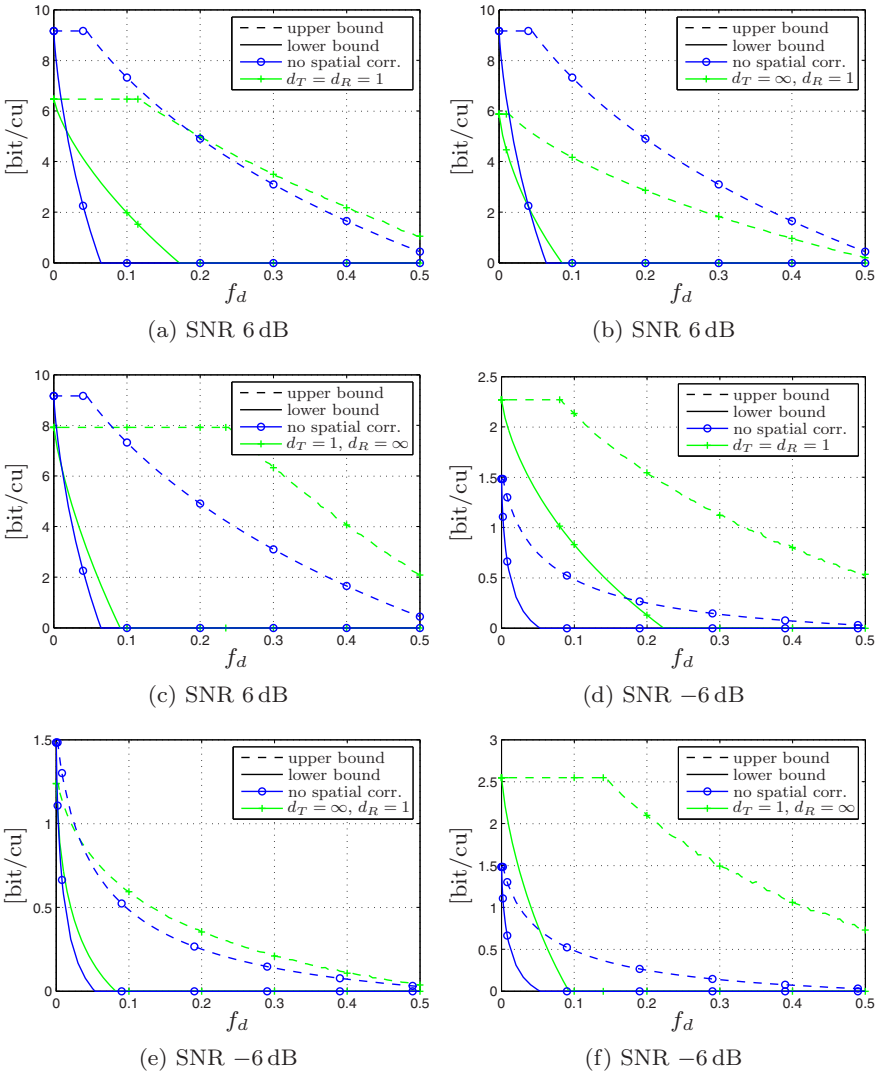
Fig. 8.4: **Effect of spatial antenna correlation on the achievable rate**, $n_T = n_R = 5$ antennas, i.i.d. zero-mean proper Gaussian input vector symbols, spatial input correlation $\mathbf{R_{\tilde{x}}}$ chosen as it would be optimal in case of perfect CSI at the receiver, $d_{T/R} = \infty$ indicates uncorrelated antennas; upper bound (8.57) in combination with coherent capacity, cf. (3.87)/(8.67); lower bound (8.65) in combination with nonnegativity of mutual information, cf. (3.94); rectangular PSD $S_h(f)$

sian input symbols. Nevertheless, it has to be stated that the upper bound in [102] holds for general PSDs of the channel fading process, whereas our bound only holds for the special case of a rectangular PSD. Furthermore, the upper bound given in [102] is a bound on capacity, whereas our upper bound is only a bound on the achievable rate, as we restrict to temporally identically distributed input symbols. This restriction has been required for our derivation. In addition, we constrain the spatial input covariance matrix and use the one that is optimal for the coherent channel.

Now, we discuss the relation between the lower bound on the peak power constrained capacity in (8.70) and the corresponding lower bound on capacity given in [102]. Therefore, note that corresponding to the SISO case, see Section 3.6, based on a time sharing argument, which corresponds to discarding the assumption on i.d. input symbols in temporal domain, the lower bound (8.70) can be enhanced, yielding

$$\sup_{\mathcal{P}_{\mathrm{MIMO}}^{\mathrm{peak}}} \mathcal{I}_L'(\mathbf{y}; \mathbf{x})$$

$$\geq \max_{\gamma \in [1,\beta]} \left\{ \frac{1}{\gamma} \mathcal{I}\left(\check{\mathbf{y}}; \check{\mathbf{x}} \middle| \tilde{\mathbf{h}}\right) \big|_{\mathrm{CM},\gamma P} - \frac{1}{\gamma} \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\gamma \frac{\psi_i p_j \phi_j}{\sigma_n^2} S_h(f) + 1\right) df \right\}$$
(8.80)

where $\mathcal{P}_{\mathrm{MIMO}}^{\mathrm{peak}}$ corresponds to $\mathcal{P}_{\mathrm{MIMO,\ i.d.}}^{\mathrm{peak}}$ in (8.37), but without the restriction to i.d. inputs in temporal domain. The bound in (8.80) corresponds in principle, i.e., except of the difference between flat and frequency selective fading to the lower bound on capacity given in [102].

8.3 Comparison with Pilot Based Synchronized Detection

As for the SISO case, we now compare the previously derived bounds on the achievable rate with an i.i.d. zero-mean proper Gaussian input distribution to bounds on the achievable rate with synchronized detection, where the channel estimation is solely based on pilot symbols. Here, we will restrict to the case of a spatially uncorrelated channel.

The achievable rate in a MIMO system based on the principle of synchronized detection and a pilot based channel estimation has been studied in [5], which is principally an extension of [6] to the MIMO case. Corresponding to the discussion for the SISO scenario in Chapter 5, we recall the results of [5] for the MIMO case in this section.

Fig. 8.5 shows a block diagram of a synchronized detection based MIMO system. As already explained in Chapter 5, the channel is estimated based on pilot symbols which are periodically inserted into the transmit symbol stream.

The channel estimates are used for a coherent, i.e., symbol-wise, detection. Due to the fact that in a MIMO system, the signal at each receive antenna is a superposition of the transmitted signals of all transmit antennas, orthogonal pilot sequences are required to enable estimation of each subchannel, i.e., from each transmit antenna j to each receive antenna i. To achieve orthogonality, n_T pilot symbols are required within one channel sampling period L, where L has to fulfill (5.8) to sample the channel fading process at least with Nyquist rate.

Based on the estimates $\hat{h}_{ij}^{\mathrm{pil}}(k)$ of the channel fading weights $h_{ij}(k)$, the channel output $\tilde{\mathbf{y}}(k)$ can be rewritten as

$$\tilde{\mathbf{y}}(k) = \hat{\tilde{\mathbf{H}}}_{\mathrm{pil}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{E}}_{\mathrm{pil}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k) \tag{8.81}$$

with $[\hat{\tilde{\mathbf{H}}}_{\mathrm{pil}}(k)]_{i,j} = \hat{h}_{ij}^{\mathrm{pil}}(k)$ and where $[\tilde{\mathbf{E}}_{\mathrm{pil}}(k)]_{ij} = e_{ij}^{\mathrm{pil}}(k) = h_{ij}(k) - \hat{h}_{ij}^{\mathrm{pil}}(k)$ is the channel estimation error.

The channel estimation error variance is given by, cf. (5.15)

$$\sigma_{e_{\mathrm{pil}}}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} df \tag{8.82}$$

which for the case of a rectangular PSD of the channel fading process yields, cf. (5.16)

$$\sigma_{e_{\mathrm{pil}}}^2 = \frac{2f_d L}{2f_d L + \rho}\sigma_h^2 \tag{8.83}$$
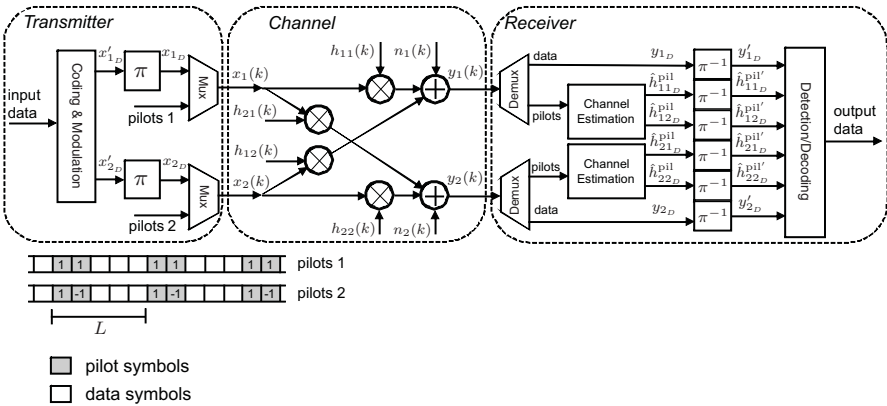
Fig. 8.5: Block diagram of a synchronized detection receiver with a pilot based channel estimation, MIMO flat-fading channel with $n_T = n_R = 2$ antennas; π/π^{-1} interleaving / deinterleaving, example of an orthogonal pilot scheme, the subscript D denotes a data symbol time instant

with ρ being the average SNR as defined in (8.13). Note, for (8.83) it is assumed that the pilot power is equal to the average data symbol power.

Obviously, this estimation error variance corresponds to the one in the SISO case, although the transmit power is now split over several antennas. The use of orthogonal pilot sequence leads to this fact. This observation can be easily verified, considering that orthogonality can be achieved by transmitting a pilot symbol with power P from one transmit antenna while the other antennas transmit nothing at this time instant. This is then repeated for all transmit antennas.

As already discussed in the SISO case, the channel estimation errors at different time instances are correlated. This temporal correlation is — due to interleaving — not exploited by the channel decoding unit. Therefore, in the following we calculate bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation based on artificial channel estimates $\breve{\hat{\mathbf{H}}}_{\mathrm{pil}}(k)$ that have the same statistics as the channel estimates $\hat{\mathbf{H}}_{\mathrm{pil}}(k)$, except that the corresponding channel estimation errors are temporally uncorrelated, see the discussion in Section 5.2.

8.3.1 Achievable Rate with Pilot Based Synchronized Detection

Corresponding to the SISO case, the achievable rate is degraded by two factors with respect to the case of perfect channel knowledge. Firstly, there is a degradation because the time instances where pilot symbols are transmitted are lost for data transmission. This degradation leads to a pre-log factor $(L - n_T)/L$. Secondly, there is a loss as the SNR is degraded due to the channel estimation error by the factor

$$\eta = \frac{\sigma_h^2 - \sigma_{e_{\mathrm{pil}}}^2}{\sigma_n^2 + \sigma_{e_{\mathrm{pil}}}^2 P} \frac{\sigma_n^2}{\sigma_h^2}. \tag{8.84}$$

In [5], it is shown that the achievable rate with synchronized detection $\mathcal{I}'\big(\mathbf{y}; \mathbf{x}\big|\breve{\hat{\mathbf{h}}}_{\mathrm{pil}}\big)$, where $\breve{\hat{\mathbf{h}}}_{\mathrm{pil}}$ is the estimate of $\mathbf{h}$, which is additionally characterized by a temporally uncorrelated estimation error, is for zero-mean proper Gaussian data symbols lower-bounded by

$$\mathcal{I}'\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) \geq \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)$$

$$= \frac{L-n_T}{L}\mathrm{E}_{\tilde{\mathbf{H}}(k)}\log\det\left(\mathbf{I}_{n_R} + \frac{P}{n_T\sigma_n^2}\eta\tilde{\mathbf{H}}(k)(\tilde{\mathbf{H}}(k))^H\right) \tag{8.85}$$

$$= \frac{L-n_T}{L}\int_0^\infty \log\left(1 + \frac{P}{n_T}\frac{\sigma_h^2-\sigma_{e_{\mathrm{pil}}}^2}{\sigma_{e_{\mathrm{pil}}}^2 P+\sigma_n^2}z\right)\sum_{k=0}^{m-1}\frac{k![L_k^{n-m}(z)]^2}{(k+n-m)!}z^{n-m}e^{-z}dz$$

$$= \frac{L-n_T}{L}\int_0^\infty \log\left(1 + \frac{\rho}{n_T}\eta z\right)\sum_{k=0}^{m-1}\frac{k!}{(k+n-m)!}[L_k^{n-m}(z)]^2 z^{n-m}e^{-z}dz \tag{8.86}$$

where the second term is the capacity of a MIMO channel with an SNR degradation η, cf. (8.64) and (8.66). Furthermore, in [5] it is shown that $\mathcal{I}'(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}})$ is upper-bounded by

$$\mathcal{I}'\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) \leq \mathcal{I}'_U\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) + \frac{L-n_T}{L}n_R\mathrm{E}_{\tilde{\mathbf{x}}(k)}\log\left(\frac{P\sigma_{e_{\mathrm{pil}}}^2+\sigma_n^2}{(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)\sigma_{e_{\mathrm{pil}}}^2+\sigma_n^2}\right) \tag{8.87}$$

where $\tilde{\mathbf{x}}(k)$ is the input vector at time instances where data symbols are transmitted, and $\mathrm{E}[(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)] = P$. The second term in (8.87) accounts for the non-Gaussianity of the term $\hat{\mathbf{E}}_{\mathrm{pil}}(k)\tilde{\mathbf{x}}(k)$ in (8.81). It is nonnegative and approaches zero for $\sigma_{e_{\mathrm{pil}}}^2 \to 0$.

For zero-mean proper Gaussian data symbols (8.87) yields

$$\mathcal{I}'_U\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)\Big|_{\mathrm{PG}}$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) + \frac{L-n_T}{L}n_R\int_{z=0}^\infty \log\left(\frac{P\sigma_{e_{\mathrm{pil}}}^2+\sigma_n^2}{\frac{P}{n_T}\sigma_{e_{\mathrm{pil}}}^2 z+\sigma_n^2}\right)\frac{z^{n_T-1}e^{-z}}{\Gamma(n_T)}dz$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) + \Delta_{\mathcal{I}(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}})}\Big|_{\mathrm{PG}}. \tag{8.88}$$

With increasing n_T, the difference $\Delta_{\mathcal{I}(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}})}\big|_{\mathrm{PG}}$ decreases. This can be intuitively seen based on the fact that $(\tilde{\mathbf{x}}(k))^H\tilde{\mathbf{x}}(k)$ in (8.87) converges to P for $n_T \to \infty$. On the other hand, $\Delta_{\mathcal{I}(\mathbf{y};\mathbf{x}|\overset{\scriptscriptstyle\smile}{\hat{\mathbf{h}}}_{\mathrm{pil}})}\big|_{\mathrm{PG}}$ increases with n_R.

Fig. 8.6 shows the achievable rate with synchronized detection in combination with a solely pilot based channel estimation in comparison to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. The pilot spacing L is chosen such that the lower bound in (8.86) is maximized. As the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs are not tight, it is not possible to give general statements on the performance of synchronized detection with a solely pilot based chan-
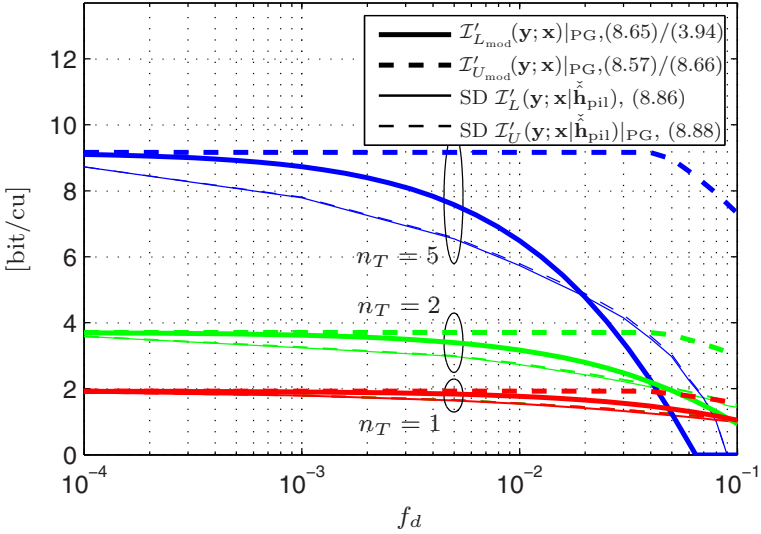
Fig. 8.6: **Comparison of bounds on the achievable rate with pilot based synchronized detection (SD) to the achievable rate with i.i.d. symbols**; in both cases i.i.d. zero-mean proper Gaussian (data) symbols are assumed, MIMO, $n_T = n_R$, SNR 6 dB; lower and upper bound for SD partially overlap; the upper and lower bound on the achievable rate with i.i.d. inputs in (8.57) and (8.65) are modified by the coherent capacity and the nonnegativity of mutual information, cf. (3.87)/(8.66) and (3.94); rectangular PSD $S_h(f)$

nel estimation. However, it can be observed that in the important region of small channel dynamics, which correspond to realistic mobile channels, the achievable rate with pilot based synchronized detection without pilot-to-data power ratio optimization stays below the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. This shows that synchronized detection with a solely pilot based channel estimation is not optimal with respect to the channel capacity.

8.3.1.1 Optimized Pilot-to-Data Power Ratio

Corresponding to the SISO case, see Section 5.2.2, the achievable rate can be increased by optimizing the ratio ν between power spent for pilot symbols and power used for data symbols, see (5.43). The achievable rate for this case has been examined in [4] and the results will be recalled briefly in the following. As in the SISO case, we use the approximation

$$L = \frac{1}{2f_d} \tag{8.89}$$

for $L = \lfloor 1/(2f_d) \rfloor$, see the discussion in Section 5.2.2.

Due to the additional pilots required to achieve orthogonality, the data power σ_d^2 and the pilot power σ_p^2 are given by (cf. (5.45) and (5.46) for the SISO case)

$$\sigma_p^2 = \frac{1}{2f_d n_T + \frac{1}{\nu}(1 - 2f_d n_T)} P \tag{8.90}$$

$$\sigma_d^2 = \frac{1}{1 - 2f_d n_T + 2f_d n_T \nu} P \tag{8.91}$$

where these powers are the sum powers transmitted over all transmit antennas at one time instant, i.e.,

$$\mathrm{E}\left[(\tilde{\mathbf{x}}_D(k))^H \tilde{\mathbf{x}}_D(k)\right] = \sigma_d^2 \tag{8.92}$$

$$(\tilde{\mathbf{x}}_P(k))^H \tilde{\mathbf{x}}_P(k) = \sigma_p^2. \tag{8.93}$$

Here the indices D and P notify data and pilot symbols.

For a rectangular PSD $S_h(f)$ and, thus, with $\sigma_{e_{\mathrm{pil}}}^2$ in (8.83) (for the SNR of the pilot symbols), and with the pilot spacing in (8.89) the SNR degradation factor η becomes

$$\eta = \frac{\rho\nu}{(\rho(1 + \nu) - 2f_d n_T(1 - \nu) + 1)(2f_d n_T(\nu - 1) + 1)}. \tag{8.94}$$

For a MIMO system, the optimum ratio ν given in (5.48) for the SISO case generalizes for the MIMO scenario to [4]

$$\nu_{\mathrm{opt}} = \sqrt{\frac{(1 - 2f_d n_T)^2 + \rho(1 - 2f_d n_T)}{2f_d n_T(2f_d n_T + \rho)}}. \tag{8.95}$$

The upper and lower bound on the achievable rate for the MIMO case are given by

$$
\begin{aligned}
\mathcal{I}'\left(\mathbf{y}; \mathbf{x} | \check{\mathbf{h}}_{\mathrm{pil}}\right) &\geq \mathcal{I}'_L\left(\mathbf{y}; \mathbf{x} | \check{\mathbf{h}}_{\mathrm{pil}}\right) \\
&= (1 - 2f_d n_T) \mathrm{E}_{\tilde{\mathbf{H}}(k)} \log\det\left(\mathbf{I}_{n_R} + \frac{P}{n_T \sigma_n^2} \eta \tilde{\mathbf{H}}(k)(\tilde{\mathbf{H}}(k))^H\right) \\
&= (1 - 2f_d n_T) \int_0^\infty \log\left(1 + \frac{\rho}{n_T} \eta z\right) \sum_{k=0}^{m-1} \frac{k![L_k^{n-m}(z)]^2}{(k + n - m)!} z^{n-m} e^{-z} dz
\end{aligned}
$$

$$\tag{8.96}$$

$$\mathcal{I}'\left(\mathbf{y};\mathbf{x}\big|\check{\hat{\mathbf{h}}}_{\text{pil}}\right) \leq \mathcal{I}'_U\left(\mathbf{y};\mathbf{x}\big|\check{\hat{\mathbf{h}}}_{\text{pil}}\right)$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\big|\check{\hat{\mathbf{h}}}_{\text{pil}}\right) + (1-2f_d n_T)n_R \mathrm{E}_{\tilde{\mathbf{x}}_D(k)}\log\left(\frac{\sigma_d^2\sigma_{e_{\text{pil}}}^2 + \sigma_n^2}{(\tilde{\mathbf{x}}_D(k))^H\tilde{\mathbf{x}}_D(k)\sigma_{e_{\text{pil}}}^2 + \sigma_n^2}\right).$$

$$(8.97)$$

Due to (8.89) the estimation error variance (8.83) becomes

$$\sigma_{e_{\text{pil}}}^2 = \frac{2f_d L}{2f_d L + \rho\frac{\sigma_p^2}{P}} = \frac{1}{1 + \rho\frac{\sigma_p^2}{P}} \tag{8.98}$$

with ρ and σ_p^2 given by (8.13) and (8.90).

For zero-mean proper Gaussian data symbol vectors $\tilde{\mathbf{x}}_D(k)$ (8.97) yields

$$\mathcal{I}'_U\left(\mathbf{y};\mathbf{x}\big|\check{\hat{\mathbf{h}}}_{\text{pil}}\right)\Big|_{\text{PG}}$$

$$= \mathcal{I}'_L\left(\mathbf{y};\mathbf{x}\big|\check{\hat{\mathbf{h}}}_{\text{pil}}\right) + (1-2f_d n_T)n_R \int_{z=0}^{\infty}\log\left(\frac{\sigma_d^2\sigma_{e_{\text{pil}}}^2 + \sigma_n^2}{\frac{\sigma_d^2}{n_T}\sigma_{e_{\text{pil}}}^2 z + \sigma_n^2}\right)\frac{z^{n_T-1}e^{-z}}{\Gamma(n_T)}dz.$$

$$(8.99)$$

Fig. 8.7 shows the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., without any pilot symbols, to the bounds on the achievable rate with synchronized detection in combination with a solely pilot symbol based channel estimation and pilot-to-average data symbol power ratio optimization (where we used the approximation $L = \frac{1}{2f_d}$). Also for the latter case the data symbols are i.i.d. zero-mean proper Gaussian distributed.

As the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs, i.e., without the use of any pilot symbols, are not tight, it is not possible to give general statements on the performance of pilot based synchronized detection in comparison to the achievable rate with i.i.d. Gaussian input symbols. For small channel dynamics, the achievable rate with pilot based synchronized detection is slightly smaller than the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. This indicates that from a capacity point of view synchronized detection with a solely pilot based channel estimation is not optimal. However, this does not allow the statement that the performance of pilot based synchronized detection is close to be optimal, as we do not know how tight the lower bound on the achievable rate with i.i.d. inputs given in (8.65) is, see Section 8.2.3.4.

It seems likely that with increasing channel dynamics the achievable rate with solely pilot based synchronized detection is significantly lower than the channel capacity due to the following reasons: The major difference between the SISO and the MIMO case is that in the MIMO case the amount of required pilot symbols while using synchronized detection does not only depend
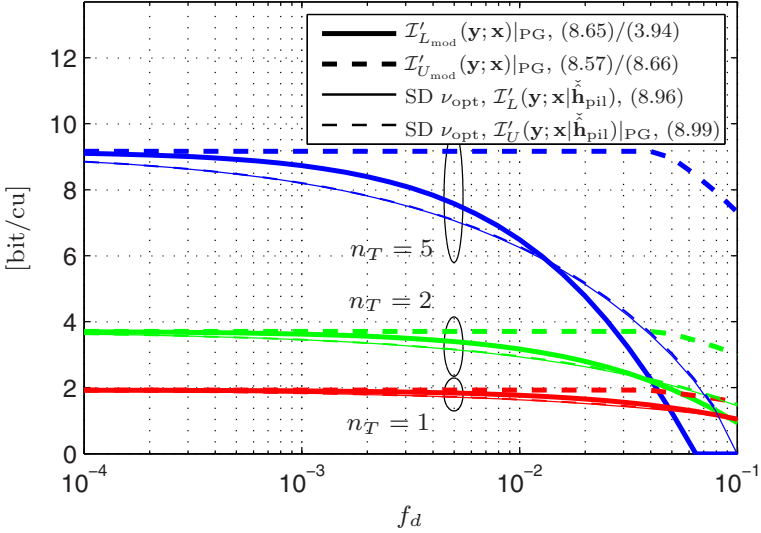
Fig. 8.7: **Comparison of bounds on the achievable rate with pilot symbol based synchronized detection (SD) and optimized pilot power to the achievable rate with i.i.d. input symbols**; in both cases i.i.d. zero-mean proper Gaussian (data) symbols are assumed; MIMO, $n_T = n_R$, SNR 6 dB; lower and upper bound for SD partially overlap; the upper and lower bound on the achievable rate with i.i.d. inputs in (8.57) and (8.65) are modified by the coherent capacity and the nonnegativity of mutual information, cf. (3.87)/(8.66) and (3.94); for SD approximation $L = \frac{1}{2f_d}$; rectangular PSD $S_h(f)$

on the channel dynamics, but also on the number of transmit antennas. As the number of required pilot symbols increases linearly with the number of transmit antennas, much space is lost for data symbols, significantly lowering the achievable rate. This loss is a constructive loss to enable the estimation of the individual subchannels based on orthogonal pilot sequences for synchronized detection. It is not given by the nature of the channel. Therefore it seems likely that in MIMO systems, with increasing channel dynamics the achievable rate with synchronized detection and a channel estimation solely based on pilot symbols stays below the channel capacity, with a gap, which increases with the number of transmit antennas and the channel dynamics. However, we have not been able to support this in general by the bounds given within this chapter. We only see that for low channel dynamics, the achievable rate with solely pilot based synchronized detection stays below the lower bound on the achievable rate with i.i.d. input symbols, showing that solely pilot based synchronized detection is not capacity-achieving.

8.4 Summary

In the present chapter, we have extended the bounds on the achievable rate
with temporally i.d. input symbols, which are given in Chapter 3 for SISO
channels, to the MIMO case, including spatial antenna correlation. We fur-
thermore assume that the input symbols on the individual transmit antennas
are zero-mean and that the spatial correlation matrix of the input symbols
over the different transmit antennas has been chosen as it would be optimal
in case of perfect receiver side channel knowledge. The main novelty is that
for the derivation of the upper bound, no peak power constraint on the in-
put symbols is required. This enables to give bounds on the achievable rate
with i.i.d. zero-mean proper Gaussian input symbols. Besides this, most of
the derivation steps used for the extension of the bounds on the achievable
rate with i.d. input symbols given in Chapter 3 to the MIMO case are sim-
ilar to [102]. Unfortunately, the derived bounds are loose, especially for an
increasing number of transmit and receive antennas.

Furthermore, we have compared the derived bounds on the achievable rate
with i.i.d. zero-mean proper Gaussian input symbols to the achievable rate
with synchronized detection in combination with a solely pilot based channel
estimation. Due to the looseness of the new bounds for i.i.d. inputs, we are
not able to give a general statement on the performance of synchronized
detection with a solely pilot based channel estimation with respect to the
channel capacity. Only for small channel dynamics we have observed that
the achievable rate with synchronized detection in combination with a solely
pilot based channel estimation stays below the achievable rate with i.i.d.
zero-mean proper Gaussian input symbols.

It should be recognized that the main bounding steps do not rely on the
assumption of a fixed spatial input signal covariance matrix $\mathbf{R}_{\tilde{x}}$. Thus, these
bounds enable a further study of the effect of the spatial input covariance
matrix.

Chapter 9
Frequency-Selective Channels

In the previous chapters, we have considered the case of frequency flat-fading channels. In this chapter, we extend the bounds on the achievable rate derived in Chapter 3 to the case of frequency-selective (wideband) channels. To get to a mathematically tractable discrete-time discrete-frequency system model, we use the approach given in [33] which is a fairly general model for an underspread wide-sense stationary uncorrelated scattering (WSSUS) scenario. For a self contained exposition, we here recall the derivation of this model, which is given in [33], and which is based on the continuous-time input-output relation of the actual channel. For a deeper discussion, we refer to [33]. The main motivation of this chapter is the derivation of bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., input symbols without a peak-power constraint. In contrast to this, the capacity bounds given in [33] are based on a peak power constraint. Therefore, both approaches mainly differ in the derivation of the upper bound, while in other places there are some similarities.

Furthermore, we recall bounds on the achievable rate with synchronized detection using a solely pilot based channel estimation in the context of frequency-selective channels and OFDM given in [4] and [47]. Additionally, we compare them with the achievable rate with i.i.d. input symbols for the case of zero-mean proper Gaussian data symbols.

As the capacity of frequency selective fading channels is not the main focus of the present work, we give a brief overview of literature in this area for further reading. In recent years there has been a wide range of publications concerning the capacity of frequency selective channels especially focussing on the large bandwidth limit. Already in the coherent case, the behavior of channel capacity in the wideband limit strongly depends on the input power constraints. I.e., for a peak power constrained input signal and infinite bandwidth the capacity can converge to zero, whereas in the case of a non-peak power constrained input the capacity converges to the capacity of an AWGN channel, which also holds for fading channels. These effects have been studied in detail in [121], [78], [128], and [117]. Most of these works also implicitly con-

sider the noncoherent case as in the wideband limit receiver knowledge of the channel fading coefficients is useless, see, e.g., [128]. Obviously, the wideband capacity strongly depends on the input signal constraints. This has also been observed in [92], where it has been shown that duty-cycled direct-sequence spread spectrum signaling is wideband capacity achieving if the number of independent fading path that can be resolved grows sublinearly with the bandwidth, whereas duty-cycled pulse position modulation only achieves the wideband capacity if the number of delay paths increases sublogarithmically. For prior work concerning the noncoherent capacity being more related to our work, we refer back to Section 1.1.

9.1 Channel Model

Up to now, we have always discussed discrete-time channel models. However, such a discrete-time channel model has to be linked to the underlying physical channel, which is time-continuous. Such a link is given in [33] in the context of the frequency selective WSSUS channel model, which is a fairly general model for a SISO channel. In this work, a solid derivation of a discrete-time discrete-frequency signal model for an underspread frequency selective WSSUS channel has been given. For completeness of presentation, we now recall this derivation.

In [33] the wireless channel is described by a linear operator $\mathbb{H}$, which projects the inputs signal $x(t)$ onto the output signal $z(t)$. The input signal $x(t)$ is an element of the Hilbert space of square integrable functions denoted by $\mathcal{L}^2$ and the output signal $z(t)$ lies in a subset of $\mathcal{L}^2$. Based on the kernel of the channel operator $k_{\mathbb{H}}(t, t')$, the noise-free input-output relation is given by

$$z(t) = \int_{t'} k_{\mathbb{H}}(t, t') x(t') dt' = \int_{\tau} h_{\mathbb{H}}(t, \tau) x(t - \tau) d\tau. \qquad (9.1)$$

For the expression at the RHS of (9.1) the variables t and t' are substituted by an absolute time t and a delay τ and, thus, $h_{\mathbb{H}}(t, \tau)$ is the *time-varying channel impulse response.*

Often the following transformations of the time-varying channel impulse response are used[1]

$$\mathcal{L}_{\mathbb{H}}(t, \nu) = \int_{\tau} h_{\mathbb{H}}(t, \tau) e^{-j2\pi\nu\tau} d\tau \qquad (9.2)$$

$$\mathfrak{S}_{\mathbb{H}}(f, \tau) = \int_{t} h_{\mathbb{H}}(t, \tau) e^{-j2\pi f t} dt. \qquad (9.3)$$

---

[1] Note, to get a consistent notation with the rest of this work, the Doppler shift is denoted by f and the frequency by ν, which is exactly the other way round as it is commonly used in literature concerning frequency selective channels.

where $\mathcal{L}_{\mathbb{H}}(t, \nu)$ is the *time-varying transfer function* and $\mathfrak{S}_{\mathbb{H}}(f, \tau)$ is named *spreading function*. Based on the latter one, the input-output relation is given by

$$z(t) = \int_f \int_\tau \mathfrak{S}_{\mathbb{H}}(f, \tau) x(t - \tau) e^{j2\pi ft} d\tau df \qquad (9.4)$$

showing that the channel output signal $z(t)$ is a superposition of weighted copies of the channel input $x(t)$. These copies are delayed in time by τ and altered in frequency by the Doppler shift f.

9.1.1 Stochastic Characterization

Often it is assumed that the time-varying transfer function $\mathcal{L}_{\mathbb{H}}(t, \nu)$ is a zero-mean jointly proper Gaussian random process in t and in ν. In [33], the publications [126] and [100] are cited, which support this assumption by measurements for narrowband and for wideband channels. As $\mathcal{L}_{\mathbb{H}}(t, \nu)$ is assumed to be zero-mean jointly proper Gaussian, its statistics are completely characterized by its correlation function, which is four-dimensional. In addition, the channel is assumed to be wide-sense stationary in time and uncorrelated in the delay domain, which is typically referred to as WSSUS [7]. Thus, $\mathcal{L}_{\mathbb{H}}(t, \nu)$ is wide-sense stationary in time and frequency, resulting in $\mathfrak{S}_{\mathbb{H}}(f, \tau)$ being uncorrelated in the Doppler shift f and the delay τ. Hence, the channel's *time-frequency correlation function* $r_{\mathbb{H}}(t, \nu)$ and the *scattering function* $S_{\mathbb{H}}(f, \tau)$, i.e., the stochastic quantities corresponding to $\mathcal{L}_{\mathbb{H}}(t, \nu)$ and $\mathfrak{S}_{\mathbb{H}}(f, \tau)$, are given by

$$r_{\mathbb{H}}(t - t', \nu - \nu') = \mathrm{E}\left[\mathcal{L}_{\mathbb{H}}(t, \nu)\mathcal{L}_{\mathbb{H}}^*(t', \nu')\right] \qquad (9.5)$$

$$S_{\mathbb{H}}(f, \tau)\delta(f - f')\delta(\tau - \tau') = \mathrm{E}\left[\mathfrak{S}_{\mathbb{H}}(f, \tau)\mathfrak{S}_{\mathbb{H}}^*(f', \tau')\right]. \qquad (9.6)$$

The two functions are related by a two-dimensional Fourier transform

$$S_{\mathbb{H}}(f, \tau) = \int_t \int_\nu r_{\mathbb{H}}(t, \nu) e^{-j2\pi(tf - \tau\nu)} dt d\nu. \qquad (9.7)$$

The function $S_{\mathbb{H}}(f, \tau)$ corresponds to the spectrum of the channel fading process.

9.1.2 The Underspread Assumption

As the velocity of the transmitter, the receiver, and of objects in the environment are limited, cf. Chapter 2, the maximum Doppler shift f_0 is limited.

Furthermore, we assume that the maximum delay is given by $2\tau_0$. To simplify notation and without loss of generality, we assume scattering functions that are centered at $\tau = 0$ and $f = 0$. This corresponds to the removal of all fixed frequency and delay shifts, which, as they are constant over time, have no influence on the capacity. Consequently, the scattering function is supported on a rectangle of spread $\Delta_\mathbb{H} = 4f_0\tau_0$, i.e.,

$$S_\mathbb{H}(f, \tau) = 0 \quad \text{for } (f, \tau) \notin [-f_0, f_0] \times [-\tau_0, \tau_0]. \tag{9.8}$$

The channel is named *underspread*, if $\Delta_\mathbb{H} < 1$ holds, see [33] and references there in. Most wireless communication channels are highly underspread, i.e., with $\Delta_\mathbb{H} \approx 10^{-3}$ for typical mobile channels, see, e.g., [79] and references in [33]. This underspread assumption will be important in the following, as this is a prerequisite to discretize the channel operator based on a well-structured set of approximate eigenfunctions.

The problem that arises in the discretization of the channel operator $\mathbb{H}$ is the fact that the left and right singular functions of its kernel are themselves random. As the set of singular functions form the basis of the input and the output space, their knowledge is required for a diagonalization of the channel operator $\mathbb{H}$. Therefore, the diagonalization of the channel, which leads to a set of scalar input-output relations, requires perfect knowledge of the channel, which is not available.

However, due to the underspread assumption, the eigenfunctions of $\mathbb{H}$ are approximately normal, allowing to approximate the singular value decomposition by an eigenvalue decomposition. Thus, any signal $g(t)$ that is well localized in time and frequency is an approximate eigenfunction of $\mathbb{H}$, i.e., the mean-squared error $\mathrm{E}\left[||\langle \mathbb{H}g, g\rangle g - \mathbb{H}g||^2\right]$ is small due to the underspread assumption. Therefore, based on a set of deterministic orthogonal eigenfunctions in [33] an approximate channel operator $\tilde{\mathbb{H}}$ is constructed. Beside the fact that these eigenfunctions need to be well localized in time and frequency, they should be well structured to be suitable for further analysis. As the support of the scattering function $S_\mathbb{H}(f, \tau)$ is strictly limited, the samples $\mathcal{L}_\mathbb{H}(nT, kF)$ with $T \leq 1/(2f_0)$ and $F \leq 1/(2\tau_0)$ completely describe $\mathcal{L}_\mathbb{H}(t, \nu)$ due to the Nyquist theorem. In [33] the *Weyl-Heisenberg set* has been chosen as a well structured set of eigenfunctions. This set is given by $\{g_{n,k}(t)\}$, with the orthonormal functions $g_{n,k}(t) = g(t - nT)e^{j2\pi kFt}$. To be well localized in time and frequency it is required that $TF > 1$, see [33] which cites [16]. A large product TF leads to an improved time-frequency localization of $g(t)$, but decreases the dimensions of the signaling space compared to a choice of TF close to one.

The approximate channel operator $\tilde{\mathbb{H}}$ can be easily diagonalized due to the deterministic set of orthonormal functions. Obviously, this requires no knowledge on the realization of $\mathbb{H}$. Furthermore, practical transmit signals $x(t)$ are bandlimited and, thus, can be modeled as a finite sum of K frequency slots due to the fact that $g(t)$ is well concentrated in frequency, i.e.,

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{K-1} \langle x, g_{n,k} \rangle g_{n,k}(t) \tag{9.9}$$

where $\langle x, g_{n,k} \rangle = x[n, k]$ is the transmit symbol at time slot n in the frequency slot k.

Using the approximate channel operator $\tilde{\mathbb{H}}$, the received signal is given by

$$y(t) = (\tilde{\mathbb{H}}x)(t) + n(t) \tag{9.10}$$

where $n(t)$ is additive white Gaussian noise. Projection of $y(t)$ onto $\{g_{n,k}(t)\}$ delivers

$$
\begin{aligned}
y[n, k] = \langle y, g_{n,k} \rangle &= \langle \tilde{\mathbb{H}}x, g_{n,k} \rangle + \langle n, g_{n,k} \rangle \\
&= \sum_{n',k'} x[n', k'] \langle \tilde{\mathbb{H}}g_{n',k'}, g_{n,k} \rangle + n[n, k] \\
&= \mathcal{L}_{\mathbb{H}}(nT, kF)x[n, k] + n[n, k] \\
&= h[n, k]x[n, k] + n[n, k].
\end{aligned}
\tag{9.11}
$$

Thus, an input-output relation for all time and frequency slots (n, k) is found. Orthonormality of the set $\{g_{n,k}(t)\}$ implies that $n[n, k]$ is jointly proper Gaussian.

9.1.2.1 OFDM Interpretation

Furthermore, in [33] the following interpretation of the discussed approximate channel operator is given. Obviously, the input signal $x(t)$ in (9.9) is given by the modulation of discrete-data symbols onto a set of orthogonal signals, which corresponds to pulse-shaped OFDM [65]. Based on this interpretation, it is possible to give a physical interpretation of the error that arises by approximating the genuine channel operator $\mathbb{H}$ by $\tilde{\mathbb{H}}$. Therefore, we project the noise free version of $y(t)$, which is named $z(t)$, onto the eigenfunctions $g_{n,k}(t)$

$$
\begin{aligned}
\langle z, g_{n,k} \rangle &= \sum_{n'=-\infty}^{\infty} \sum_{k'=0}^{K-1} x[n', k'] \langle \mathbb{H}g_{n',k'}, g_{n,k} \rangle \\
&= \langle \mathbb{H}g_{n,k}, g_{n,k} \rangle x[n, k] + \sum_{\substack{n'=-\infty \\ (n',k') \neq (n,k)}}^{\infty} \sum_{k'=0}^{K-1} x[n', k'] \langle \mathbb{H}g_{n',k'}, g_{n,k} \rangle
\end{aligned}
\tag{9.12}
$$

where the first term on the RHS (9.12) is the desired signal and the second term arises due to intersymbol interference (ISI) and intercarrier interference (ICI). For the approximative channel $\tilde{\mathbb{H}}$ the first term on the RHS of (9.12)

is approximated by $\mathcal{L}_{\mathbb{H}}(nT, kF)x[n, k]$. Thus, using (9.11) can be identified as a pulse shaped OFDM transmission over the genuine channel $\mathbb{H}$ ignoring ISI and ICI.

If the function $g(t)$ is well chosen, ISI and ICI can be reduced. This reduction increases with an increasing product TF, but leads to a loss in dimensions of the signaling space. A factor of $TF \simeq 1.25$ is given as a good tradeoff between both effects in [33], which in this regard refers, e.g., to [65] and [76]. A corresponding loss of signaling dimensions also occurs due to the cyclic prefix when using cyclic prefix OFDM.

9.1.3 Discrete-Time Discrete-Frequency Input-Output Relation

Based on the approximative channel operator $\tilde{\mathbb{H}}$ [33] gives a discrete-time discrete-frequency input-output relation of the frequency-selective fading channel. With (9.11) the input-output relation for each individual time and frequency slot is given. Concatenating the input-output relations in frequency dimension leads to the following matrix-vector notation

$$\tilde{\mathbf{y}}_n = \tilde{\mathbf{H}}_n \tilde{\mathbf{x}}_n + \tilde{\mathbf{n}}_n \tag{9.13}$$

where n indicates the n-th time slot. The input vector $\tilde{\mathbf{x}}_n$, the output vector $\tilde{\mathbf{y}}_n$, the noise vector $\tilde{\mathbf{n}}_n$ and the channel matrix $\tilde{\mathbf{H}}_n$ all at time slot n are defined as follows

$$\tilde{\mathbf{x}}_n = [x[n, 0], \dots, x[n, K-1]]^T \in \mathbb{C}^K \tag{9.14}$$

$$\tilde{\mathbf{y}}_n = [y[n, 0], \dots, y[n, K-1]]^T \in \mathbb{C}^K \tag{9.15}$$

$$\tilde{\mathbf{n}}_n = [n[n, 0], \dots, n[n, K-1]]^T \in \mathbb{C}^K \tag{9.16}$$

$$\tilde{\mathbf{h}}_n = [h[n, 0], \dots, h[n, K-1]]^T \in \mathbb{C}^K \tag{9.17}$$

$$\tilde{\mathbf{H}}_n = \mathrm{diag}\left(\tilde{\mathbf{h}}_n\right) \tag{9.18}$$

where K is the number of used subcarriers. Notice, here the tilde is used for the subcarrier domain, in contrast to Chapter 8 where it is used for the spatial domain.

Further concatenation over time domain leads to the input-output relation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{9.19}$$

where the vectors are defined by

$$\mathbf{x} = \left[\tilde{\mathbf{x}}_1^T, \ldots, \tilde{\mathbf{x}}_N^T\right]^T \in \mathbb{C}^{KN} \tag{9.20}$$

$$\mathbf{y} = \left[\tilde{\mathbf{y}}_1^T, \ldots, \tilde{\mathbf{y}}_N^T\right]^T \in \mathbb{C}^{KN} \tag{9.21}$$

$$\mathbf{n} = \left[\tilde{\mathbf{n}}_1^T, \ldots, \tilde{\mathbf{n}}_N^T\right]^T \in \mathbb{C}^{KN} \tag{9.22}$$

$$\mathbf{h} = \left[\tilde{\mathbf{h}}_1^T, \ldots, \tilde{\mathbf{h}}_N^T\right]^T \in \mathbb{C}^{KN} \tag{9.23}$$

$$\mathbf{H} = \mathrm{diag}(\mathbf{h}) \tag{9.24}$$

$$\mathbf{X} = \mathrm{diag}(\mathbf{x}) \tag{9.25}$$

where N is the number of considered time slots. Later on, for the derivation of the achievable rate, we will study the case $N \to \infty$.

9.1.3.1 Stochastic Characteristics

The channel correlation over the subcarrier domain and between two time slots with a temporal distance nT is given by

$$
\begin{aligned}
\mathbf{R}_{\tilde{\mathbf{h}}}[n] &= \mathrm{E}\left[\tilde{\mathbf{h}}_{n+n'}\tilde{\mathbf{h}}_{n'}^H\right] \\
&= \begin{pmatrix}
r_{\mathbb{H}}[n,0] & r_{\mathbb{H}}^*[n,1] & \ldots & r_{\mathbb{H}}^*[n,K-1] \\
r_{\mathbb{H}}[n,1] & r_{\mathbb{H}}[n,0] & \ldots & r_{\mathbb{H}}^*[n,K-2] \\
\vdots & \vdots & \ddots & \vdots \\
r_{\mathbb{H}}[n,K-1] & r_{\mathbb{H}}[n,K-2] & \ldots & r_{\mathbb{H}}[n,0]
\end{pmatrix}
\end{aligned} \tag{9.26}
$$

where the channel correlation function $r_{\mathbb{H}}[n,k]$ is given by

$$
\begin{aligned}
r_{\mathbb{H}}[n,k] &= \mathrm{E}\left[h[n+n', k+k']h^*[n',k']\right] \\
&= \mathrm{E}\left[\mathcal{L}_{\mathbb{H}}((n+n')T, (k+k')F)\mathcal{L}_{\mathbb{H}}^*(n'T, k'F)\right]. \tag{9.27}
\end{aligned}
$$

The corresponding two-dimensional PSD is given by

$$S_h(\tilde{f}, \tilde{\tau}) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} r_{\mathbb{H}}[n,k] e^{-j2\pi(n\tilde{f}-k\tilde{\tau})}, \quad |\tilde{f}|, |\tilde{\tau}| \leq 0.5 \tag{9.28}$$

where

$$\tilde{f} = fT \tag{9.29}$$

$$\tilde{\tau} = \tau F. \tag{9.30}$$

The relation between the PSD $S_h(\tilde{f}, \tilde{\tau})$ and the scattering function $S_{\mathbb{H}}(f, \tau)$ is given by [33]

$$S_h(\tilde{f}, \tilde{\tau}) = \frac{1}{TF} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} S_{\mathbb{H}}\left(\frac{\tilde{f}-n}{T}, \frac{\tilde{\tau}-k}{F}\right). \tag{9.31}$$

The variance of each channel coefficient is given by

$$\sigma_h^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(\tilde{f}, \tilde{\tau}) d\tilde{f} d\tilde{\tau} = \int_f \int_\tau S_{\mathbb{H}}(f, \tau) df d\tau \tag{9.32}$$

as the Nyquist condition holds due to $T \le 1/(2f_0)$ and $F \le 1/(2\tau_0)$.

The correlation matrix of the stacked vector $\mathbf{h}$ is given by

$$\begin{aligned}
\mathbf{R_h} &= \mathrm{E}\left[\mathbf{hh}^H\right] \\
&= \begin{pmatrix}
\mathbf{R_{\tilde{h}}}[0] & \mathbf{R_{\tilde{h}}}^H[1] & \dots & \mathbf{R_{\tilde{h}}}^H[N-1] \\
\mathbf{R_{\tilde{h}}}[1] & \mathbf{R_{\tilde{h}}}[0] & \dots & \mathbf{R_{\tilde{h}}}^H[N-2] \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{R_{\tilde{h}}}[N-1] & \mathbf{R_{\tilde{h}}}[N-2] & \dots & \mathbf{R_{\tilde{h}}}[0]
\end{pmatrix}
\end{aligned} \tag{9.33}$$

which is a two-level Hermitian Toeplitz matrix as the process $\{h[n,k]\}$ is stationary in time and frequency.

As the noise over time slots and subcarriers is independent we get

$$\mathrm{E}\left[\mathbf{nn}^H\right] = \sigma_n^2 \mathbf{I}_{KN}. \tag{9.34}$$

This completes the derivation of the discrete-time discrete-frequency system model given in [33].

As in Chapter 3 for the temporal domain, we here assume that the input signal consists of i.d. symbols in temporal and in subcarrier domain, each with an average transmit power of

$$\mathrm{E}\left[|x[n,k]|^2\right] = \alpha\sigma_x^2 \quad \forall n, k \tag{9.35}$$

with $\alpha \in [0, 1]$. Here σ_x^2 is the maximum average power per subcarrier and time slot and is given by

$$\sigma_x^2 = \frac{PT}{K} \tag{9.36}$$

with P being the maximum average transmit power.

Furthermore, the nominal mean SNR ρ is given by

$$\rho = \frac{P\sigma_h^2}{\sigma_n^2}. \tag{9.37}$$

Notice that ρ only corresponds to the average SNR if P is the actual average transmit power, i.e., $\alpha = 1$ in (9.35). As in the case of non-peak power constrained input symbols, the achievable rate is maximized by using the

maximum average transmit power, for non-peak power constrained inputs ρ corresponds to the actual mean SNR.

9.2 Bounds on the Achievable Rate

As in Chapter 3 for the flat-fading SISO case and also in Chapter 8.2 for the MIMO case, we derive bounds on the achievable rate with i.d. input symbols in temporal and in subcarrier domain. Therefore we use the separation of the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ into the channel output entropy rate $h'(\mathbf{y})$ and the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$, cf. (3.10). As in the flat-fading case discussed in Chapter 3, the restriction to i.d. input symbols is required in the following derivation.

Parts of the derivations in this section are closely related to the calculation of bounds on the capacity of frequency-selective wideband channels discussed in [33] for peak power constraint inputs. The differences between the approach taken in [33] and the approach taken in the present work will be discussed in Section 9.2.3.5.

In accordance with Chapter 3 and Chapter 8.1, we do not calculate the capacity, which is given by

$$C = \sup_{\mathcal{P}_{\text{OFDM}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty} \frac{1}{NT} \sup_{\mathcal{P}_{\text{OFDM}}} \mathcal{I}(\mathbf{y};\mathbf{x}) \qquad (9.38)$$

with

$$\mathcal{P}_{\text{OFDM}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^{NK}, \frac{1}{NT}\mathrm{E}[\mathbf{x}^H\mathbf{x}] \leq P \right\}. \qquad (9.39)$$

Instead, we restrict to input symbols which are i.d. in the temporal and in the frequency domain and therefore use the term *achievable rate* which is then given by

$$R = \sup_{\mathcal{P}_{\text{OFDM, i.d.}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \qquad (9.40)$$

where the set of input distributions is restricted to

$$\mathcal{P}_{\text{OFDM, i.d.}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^{NK}, \; p(x[n,k]) = p(x[m,l]) \; \forall k,l,n,m, \right.$$
$$\left. \mathrm{E}[|x[n,k]|^2] \leq \sigma_x^2 \; \forall n,k \right\}. \qquad (9.41)$$

In some cases, we further constrain the input using a peak power constraint. Therefore, we define

$$\mathcal{P}^{\text{peak}}_{\text{OFDM, i.d.}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^{NK}, \; p(x[n,k]) = p(x[m,l]) \; \forall k, l, n, m, \right.$$

$$\left. \left\{ \text{E}[|x[n,k]|^2] \leq \sigma_x^2, \; p(x[n,k]) = 0 \; \forall x[n,k] : |x[n,k]|^2 > \frac{P_{\text{peak}}}{K} \right\} \; \forall n, k \right\}$$

$$(9.42)$$

which effectively corresponds to the set $\mathcal{P}_{\text{OFDM, i.d.}}$ as defined in (9.41) with the further restriction of the input power for the transmit symbols on the individual subcarriers to a maximum power of P_{peak}/K, i.e., $|x[n,k]|^2 \leq P_{\text{peak}}/K$.[2] Notice, we use a peak power constraint per subcarrier. As the power is defined in the time domain, i.e., based on a complete OFDM symbol, cf. (9.36), we use the notation P_{peak}/K for the peak power per subcarrier. Furthermore, recognize that the peak power constraint per subcarrier is not only a mathematical requirement, sometimes regulation also imposes peak power constraints in specific frequency bands.

9.2.1 The Channel Output Entropy Rate $h'(\mathbf{y})$

9.2.1.1 Lower Bound on $h'(\mathbf{y})$

Analogous to Section 3.2.1 for the flat-fading SISO scenario, the channel output entropy rate $h'(\mathbf{y})$ is lower bounded by, cf. (3.13)

$$h'(\mathbf{y}) \geq \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) \tag{9.43}$$

where the first term in (9.43) is the coherent mutual information rate. It will be further evaluated in Section 9.2.3.2 depending on the input constraints.

The second term on the RHS of (9.43) arises due to the additive white Gaussian noise and, thus, is given by

$$h'(\mathbf{y}|\mathbf{x}, \mathbf{h}) = \frac{1}{NT} \log \det \left(\pi e \sigma_n^2 \mathbf{I}_{KN} \right) = \frac{K}{T} \log \left(\pi e \sigma_n^2 \right). \tag{9.44}$$

9.2.1.2 Upper Bound on $h'(\mathbf{y})$

Corresponding to Section 3.2.2, we now derive an upper bound on $h'(\mathbf{y})$. Analogously to (3.16), we use the fact that zero-mean jointly proper Gaussian random variables are entropy maximizers for a given correlation matrix, thus

$$h(\mathbf{y}) \leq \log \det \left(\pi e \mathbf{R_y} \right). \tag{9.45}$$

---

[2] Note that differently to P, P_{peak} is normalized by the OFDM symbol duration T.

Assuming an average transmit power per subcarrier of $\alpha\sigma_x^2$ with $\alpha \in [0,1]$, the diagonal elements of the correlation matrix $\mathbf{R_y}$ are given by

$$[\mathbf{R_y}]_{l,l} = \mathrm{E}\left[|y[l,l]|^2\right] = \alpha\sigma_x^2\sigma_h^2 + \sigma_n^2. \tag{9.46}$$

Thus, using Hadamard's inequality the entropy rate $h'(\mathbf{y})$ is upper-bounded by

$$
\begin{aligned}
h'(\mathbf{y}) &\leq \frac{1}{NT}\log\det\left(\pi e\left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)\mathbf{I}_{KN}\right)\\
&= \frac{K}{T}\log\left(\pi e\left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)\right) = h'_U(\mathbf{y}). \tag{9.47}
\end{aligned}
$$

9.2.2 The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$

Corresponding to Section 3.3, we derive an upper and a lower bound on the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ within this section.

9.2.2.1 Upper Bound on $h'(\mathbf{y}|\mathbf{x})$

Analogous to Section 3.3.1, the PDF of the conditional channel output is zero-mean proper Gaussian with the correlation matrix notationally corresponding to (3.20). Thus, as in the flat-fading case, the conditional channel output entropy $h(\mathbf{y}|\mathbf{x})$ can be upper-bounded using Jensen's inequality and the concavity of the log-function by, cf. (3.24)

$$
\begin{aligned}
h(\mathbf{y}|\mathbf{x}) &= \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e\left(\mathbf{X}\mathbf{R_h}\mathbf{X}^H + \sigma_n^2\mathbf{I}_{KN}\right)\right)\right] \tag{9.48}\\
&\overset{(a)}{=} \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e\left(\mathbf{X}^H\mathbf{X}\mathbf{R_h} + \sigma_n^2\mathbf{I}_{KN}\right)\right)\right]\\
&\overset{(b)}{\leq} \log\det\left(\pi e\left(\alpha\sigma_x^2\mathbf{R_h} + \sigma_n^2\mathbf{I}_{KN}\right)\right)\\
&= \log\det\left(\frac{\alpha\sigma_x^2}{\sigma_n^2}\mathbf{R_h} + \mathbf{I}_{KN}\right) + KN\log\left(\pi e\sigma_n^2\right) \tag{9.49}
\end{aligned}
$$

where for (a) we have used the identity given in (3.25) and (b) follows from the concavity of log det on the set of positive definite matrices. Furthermore, for (b) we assume that the average transmit power per subcarrier is given by $\alpha\sigma_x^2$ with $\alpha \in [0,1]$ to allow for average transmit powers smaller than σ_x^2.

For the evaluation of the achievable rate, we study the case of an infinite long transmission interval, i.e., $N \to \infty$. For this case the entropy rate $h'(\mathbf{y}|\mathbf{x})$ is upper-bounded by

$$h'(\mathbf{y}|\mathbf{x}) \leq \lim_{N \to \infty} \frac{1}{NT} \log \det \left(\frac{\alpha \sigma_x^2}{\sigma_n^2} \mathbf{R_h} + \mathbf{I}_{KN} \right) + \frac{K}{T} \log \left(\pi e \sigma_n^2 \right). \qquad (9.50)$$

Now the derivation differs from the flat-fading case as the matrix $\mathbf{R_h}$ is a two-level Toeplitz matrix. Based on an extension of Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices given in [80, Theorem 3.4], [39], which has been also applied on equal respectively similar problems in [33] and [96], the RHS (9.50) is given by

$$h'(\mathbf{y}|\mathbf{x}) \leq \lim_{N \to \infty} \frac{1}{NT} \log \det \left(\frac{\alpha PT}{\sigma_n^2 K} \mathbf{R_h} + \mathbf{I}_{KN} \right) + \frac{K}{T} \log \left(\pi e \sigma_n^2 \right)$$

$$= \frac{1}{T} \int_{\tilde{f}=-\frac{1}{2}}^{\frac{1}{2}} \log \det \left(\frac{\alpha PT}{\sigma_n^2 K} \mathbf{S}_h(\tilde{f}) + \mathbf{I}_K \right) d\tilde{f} + \frac{K}{T} \log \left(\pi e \sigma_n^2 \right) \quad (9.51)$$

where

$$\mathbf{S}_h(\tilde{f}) = \sum_{n=-\infty}^{\infty} \mathbf{R}_{\tilde{\mathbf{h}}}[n] e^{-j2\pi n \tilde{f}}, \quad |\tilde{f}| \leq 0.5 \qquad (9.52)$$

is the matrix valued power spectral density of the multivariate channel fading process $\{\tilde{\mathbf{h}}_n\}$.

Opposed to the time domain where we have been interested in the limit of an infinitely long transmission interval and thus we could apply Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices, the number of subcarriers in the frequency domain K is finite. Thus, the evaluation of (9.51) requires the calculation of the eigenvalues of $\mathbf{S}_h(\tilde{f})$. As there is no closed form solution for the eigenvalues of $\mathbf{S}_h(\tilde{f})$, this requires a numerical evaluation which will be challenging for large bandwidths, i.e., a large number of subcarriers K. In [33, Lemma 3], the following upper bound on first term on the RHS of (9.51) is given

$$\frac{1}{T} \int_{\tilde{f}=-\frac{1}{2}}^{\frac{1}{2}} \log \det \left(\frac{\alpha PT}{\sigma_n^2 K} \mathbf{S}_h(\tilde{f}) + \mathbf{I}_K \right) d\tilde{f} \leq 2 f_0 \sum_{i=0}^{K-1} \log \left(1 + \frac{\alpha P}{\sigma_n^2 2 f_0 K} d_i \right)$$
$$(9.53)$$

where

$$d_i = \text{Re} \left\{ \frac{2}{K} \sum_{k=0}^{K-1} (K - k) r_{\mathbb{H}}[0, k] e^{-j2\pi \frac{ik}{K}} \right\} - r_{\mathbb{H}}[0, 0]. \qquad (9.54)$$

In Appendix A.13, we recall the proof of inequality (9.53) given in [33].

With (9.53) and (9.51), we get the following upper bound on $h'(\mathbf{y}|\mathbf{x})$

$$h'(\mathbf{y}|\mathbf{x}) \leq 2f_0 \sum_{i=0}^{K-1} \log\left(1 + \frac{\alpha P}{\sigma_n^2 2 f_0 K} d_i\right) + \frac{K}{T} \log\left(\pi e \sigma_n^2\right)$$

$$= h'_U(\mathbf{y}|\mathbf{x}). \tag{9.55}$$

9.2.2.2 Lower Bound on $h'(\mathbf{y}|\mathbf{x})$ for a Brick-Shaped Scattering Function

In this section, we extend the lower bound on the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$, which has been derived in Section 3.3.2 for the flat-fading SISO case, to the frequency-selective case. As in the flat-fading case, we have to make some restrictions to the scattering function $S_{\mathbb{H}}(f, \tau)$. We assume that the scattering function is brick-shaped, i.e., that its support region is given by $(f, \tau) \in ([-f_0, f_0] \times [-\tau_0, \tau_0])$ and that within this interval $S_{\mathbb{H}}(f, \tau)$ is constant. Thus, we assume

$$S_h(\tilde{f}, \tilde{\tau}) = \begin{cases} \frac{\sigma_h^2}{4\tilde{f}_0 \tilde{\tau}_0} & \left(\tilde{f}, \tilde{\tau}\right) \in [-\tilde{f}_0, \tilde{f}_0] \times [-\tilde{\tau}_0, \tilde{\tau}_0] \\ 0 & (\tilde{f}_0 < |\tilde{f}| < 0.5) \vee (\tilde{\tau}_0 < |\tilde{\tau}| < 0.5) \end{cases}. \tag{9.56}$$

For the evaluation of the lower bound, we introduce the following stacking of the channel fading weights over time. We define the vector $\bar{\mathbf{h}}_k$ containing all channel fading weights over time for the subcarrier k, i.e.,

$$\bar{\mathbf{h}}_k = [h[1, k], \ldots, h[N, k]]^T \in \mathbb{C}^N. \tag{9.57}$$

The autocorrelation matrix for the temporal domain is given by

$$\mathbf{R}_{\bar{\mathbf{h}}}[k] = \mathrm{E}\left[\bar{\mathbf{h}}_{k+k'}\bar{\mathbf{h}}_{k'}^H\right]$$
$$= \begin{pmatrix} r_{\mathbb{H}}[0, k] & r_{\mathbb{H}}^*[1, k] & \ldots r_{\mathbb{H}}^*[N-1, k] \\ r_{\mathbb{H}}[1, k] & r_{\mathbb{H}}[0, k] & \ldots r_{\mathbb{H}}^*[N-2, k] \\ \vdots & \vdots & \ddots & \vdots \\ r_{\mathbb{H}}[N-1, k] & r_{\mathbb{H}}[N-2, k] & \ldots & r_{\mathbb{H}}[0, k] \end{pmatrix}. \tag{9.58}$$

Due to our assumption on a brick-shaped scattering function, the channel correlation matrix $\mathbf{R}_{\mathbf{h}}$ can be rewritten using the Kronecker product

$$\mathbf{R}_{\mathbf{h}} = \mathbf{R}_{\bar{\mathbf{h}}}[0] \otimes \mathbf{R}_{\tilde{\mathbf{h}}}[0] \tag{9.59}$$

with $\mathbf{R}_{\tilde{\mathbf{h}}}[0]$ as defined in (9.26) and $\mathbf{R}_{\bar{\mathbf{h}}}[0]$ as defined in (9.58).

Proceeding along the lines of the proof for the flat-fading case given in Section 3.3.2, based on the asymptotic equivalence of Toeplitz and circulant matrices, we substitute the temporal correlation matrix $\mathbf{R}_{\bar{\mathbf{h}}}[0]$ by a circulant matrix $\mathbf{C}_{\bar{\mathbf{h}}}[0]$ with the decreasingly ordered eigenvalues given by

$$\lambda_{\bar{\mathbf{h}},n} = \begin{cases} \frac{1}{2\tilde{f}_0} & \text{for } 1 \leq n \leq 2\tilde{f}_0 N \\ 0 & \text{otherwise} \end{cases} . \qquad (9.60)$$

Due to the asymptotic equivalence of Toeplitz and circulant matrices, for $N \to \infty$, this corresponds to a rectangular shaped PSD of the channel fading process. Therefore, to calculate the entropy rate $h'(\mathbf{y}|\mathbf{x})$ we substitute $\mathbf{R_h}$ by

$$\mathbf{C_h} = \mathbf{C}_{\bar{\mathbf{h}}}[0] \otimes \mathbf{R}_{\tilde{\mathbf{h}}}[0]. \qquad (9.61)$$

Furthermore, we use the spectral decomposition of $\mathbf{C_h}$ given by

$$\mathbf{C_h} = \mathbf{U} \Lambda_{\mathbf{h}} \mathbf{U}^H \qquad (9.62)$$

where the diagonal matrix $\Lambda_{\mathbf{h}} = \text{diag}(\lambda_1, \ldots, \lambda_{KN})$ contains the eigenvalues of $\mathbf{C_h}$ and the matrix $\mathbf{U}$ is unitary. As the matrix $\mathbf{C_h}$ can be expressed as a Kronecker product of $\mathbf{C}_{\bar{\mathbf{h}}}[0]$ and $\mathbf{R}_{\tilde{\mathbf{h}}}[0]$, its eigenvalues are given by

$$\left\{ \lambda_i = \lambda_{\bar{\mathbf{h}},n} \lambda_{\tilde{\mathbf{h}},k} \,\middle|\, k = 1, \ldots, K, n = 1, \ldots, N \right\} \qquad (9.63)$$

with $\lambda_{\bar{\mathbf{h}},n}, n \in \{1, \ldots, N\}$ being the decreasingly ordered eigenvalues of $\mathbf{C}_{\bar{\mathbf{h}}}[0]$ and $\lambda_{\tilde{\mathbf{h}},k}, k \in \{1, \ldots, K\}$ being the eigenvalues of $\mathbf{R}_{\tilde{\mathbf{h}}}[0]$ [89].

Corresponding to (9.48) the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ is given by

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N \to \infty} \frac{1}{NT} \mathbb{E}_{\mathbf{x}} \left[\log \det \left(\pi e \left(\mathbf{X} \mathbf{R_h} \mathbf{X}^H + \sigma_n^2 \mathbf{I}_{KN} \right) \right) \right] \\ &\overset{(a)}{=} \lim_{N \to \infty} \frac{1}{NT} \left[\mathbb{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{R_h} \mathbf{X}^H \mathbf{X} + \mathbf{I}_{KN} \right) \right] + KN \log \left(\pi e \sigma_n^2 \right) \right] \end{aligned}$$
$$(9.64)$$

where for (a) we have used (3.25).

The matrix $\mathbf{R_h}$ is a two-level Toeplitz matrix, cf. (9.59). With [14, Theorem 1], which is an extension of Szegö's theorem, it can be shown that the matrix $\mathbf{R_h}$ in (9.59) is asymptotic equivalent to the matrix $\mathbf{C}_h$ in (9.61). Based on this, corresponding to Section 3.3.2, it holds that the matrix $\frac{1}{\sigma_n^2} \mathbf{R_h} \mathbf{X}^H \mathbf{X} + \mathbf{I}_{KN}$ is asymptotically equivalent to the matrix $\frac{1}{\sigma_n^2} \mathbf{C_h} \mathbf{X}^H \mathbf{X} + \mathbf{I}_{KN}$. Thus the entropy rate $h'(\mathbf{y}|\mathbf{x})$ in (9.64) can be expressed by

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N \to \infty} \frac{1}{NT} \left[\mathbb{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{C_h} \mathbf{X}^H \mathbf{X} + \mathbf{I}_{KN} \right) \right] + KN \log \left(\pi e \sigma_n^2 \right) \right] \\ &\overset{(c)}{=} \lim_{N \to \infty} \frac{1}{NT} \left[\mathbb{E}_{\mathbf{x}} \left[\log \det \left(\frac{1}{\sigma_n^2} \Lambda_{\mathbf{h}} \mathbf{U}^H \mathbf{X}^H \mathbf{X} \mathbf{U} + \mathbf{I}_{KN} \right) \right] + KN \log \left(\pi e \sigma_n^2 \right) \right] \end{aligned}$$
$$(9.65)$$

where for (c) we have used (9.62) and again (3.25).

Obviously, with (9.60) and (9.63) at maximum $\lfloor 2\tilde{f}_0 N \rfloor K$ eigenvalues of $\mathbf{C_h}$ are different from zero.

Now, we introduce a further restriction. We assume that all the eigenvalues λ_i of $\mathbf{C_h}$ that are unequal to zero have the same value c. To fulfill the power normalization given in (9.32), this constant c is given by

$$c = \frac{\sigma_h^2}{4\tilde{f}_0 \tilde{\tau}_0}. \tag{9.66}$$

This corresponds to the case that the decreasingly ordered eigenvalues of the channel correlation matrix in the frequency domain $\lambda_{\tilde{\mathbf{h}},k}$ are given by

$$\lambda_{\tilde{\mathbf{h}},k} = \begin{cases} \frac{\sigma_h^2}{2\tilde{\tau}_0} & \text{for } 1 \leq k \leq 2\tilde{\tau}_0 \cdot K \\ 0 & \text{otherwise} \end{cases}. \tag{9.67}$$

In case of infinitely many subcarriers, i.e., an infinite bandwidth, this in combination with (9.60) would correspond to the brick-shaped scattering function as given in (9.56). Obviously, in reality, the number of subcarriers K is finite, different to the transmission length N which we finally always assume to be infinite. Therefore, the assumption on the eigenvalue distribution in (9.67) can only be understood as an approximation.

Thus, by construction, $\lfloor 2\tilde{f}_0 N \rfloor \lfloor 2\tilde{\tau}_0 K \rfloor$ of the eigenvalues of the set $\{\lambda_1, \ldots, \lambda_{KN}\}$ are nonzero and all of them have the value $\frac{\sigma_h^2}{4\tilde{f}_0 \tilde{\tau}_0}$. Based on these assumptions, we can rewrite the term in the outer brackets on the RHS of (9.65) as follows

$$\mathrm{E_x}\left[\log\det\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}\tilde{\mathbf{U}}^H\mathbf{X}^H\mathbf{X}\tilde{\mathbf{U}} + \mathbf{I}_{\lfloor 2\tilde{f}_0 N \rfloor \lfloor 2\tilde{\tau}_0 K \rfloor}\right)\right] + KN\log\left(\pi e \sigma_n^2\right) \tag{9.68}$$

where the matrix $\tilde{\mathbf{U}}$ is given by

$$\tilde{\mathbf{U}} = \left[\mathbf{u}_{a_1}, \ldots, \mathbf{u}_{a_{\lfloor 2\tilde{f}_0 N \rfloor \lfloor 2\tilde{\tau}_0 K \rfloor}}\right] \in \mathbb{C}^{NK \times \lfloor 2\tilde{f}_0 N \rfloor \lfloor 2\tilde{\tau}_0 K \rfloor}. \tag{9.69}$$

Here, $\mathbf{u}_{a_i}$ are the orthonormal columns of the unitary matrix $\mathbf{U}$ with a_i taken from the set $\left\{a_1, \ldots, a_{\lfloor 2\tilde{f}_0 N \rfloor \lfloor 2\tilde{\tau}_0 K \rfloor}\right\}$ corresponding to the eigenvalues λ_i which are nonzero. Based on the inequality (3.58) in Lemma 3.1 in Section 3.3.2, we can lower-bound (9.68) as follows

$$\mathrm{E}_{\mathbf{x}}\left[\log\det\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}\tilde{\mathbf{U}}^H\mathbf{X}^H\mathbf{X}\tilde{\mathbf{U}}+\mathbf{I}_{\lfloor 2\tilde{f}_0 N\rfloor\lfloor 2\tilde{\tau}_0 K\rfloor}\right)\right]+KN\log\left(\pi e\sigma_n^2\right)$$

$$\geq\mathrm{E}_{\mathbf{x}}\left[\mathrm{trace}\left[\tilde{\mathbf{U}}^H\mathrm{diag}\left(\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x[1,1]|^2+1\right),\ldots,\right.\right.\right.$$

$$\left.\left.\left.\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x[N,K]|^2+1\right)\right)\tilde{\mathbf{U}}\right]\right]+KN\log\left(\pi e\sigma_n^2\right)$$

$$=\mathrm{trace}\left[\tilde{\mathbf{U}}^H\mathrm{diag}\left(\mathrm{E}_x\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x[1,1]|^2+1\right),\ldots,\right.\right.$$

$$\left.\left.\mathrm{E}_x\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x[N,K]|^2+1\right)\right)\tilde{\mathbf{U}}\right]+KN\log\left(\pi e\sigma_n^2\right)$$

$$\overset{(a)}{=}\sum_{i=1}^{\lfloor 2\tilde{f}_0 N\rfloor\lfloor 2\tilde{\tau}_0 K\rfloor}\mathrm{E}_x\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x|^2+1\right)+KN\log\left(\pi e\sigma_n^2\right) \qquad (9.70)$$

where for (a) we have used the fact that all $x[n,k]$ are identically distributed and that the columns of $\tilde{\mathbf{U}}$ are orthonormal.

Thus based on the approximation of the eigenvalues of $\mathbf{R}_{\tilde{\mathbf{h}}}[0]$ by the eigenvalues given in (9.67), we get the following approximative lower bound on the entropy rate

$$h'(\mathbf{y}|\mathbf{x})=\lim_{N\to\infty}\frac{1}{NT}h(\mathbf{y}|\mathbf{x})$$

$$\gtrsim\lim_{N\to\infty}\frac{1}{NT}\left\{\sum_{i=1}^{\lfloor 2\tilde{f}_0 N\rfloor\lfloor 2\tilde{\tau}_0 K\rfloor}\mathrm{E}_x\log\left(\frac{\sigma_h^2|x|^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}+1\right)+KN\log\left(\pi e\sigma_n^2\right)\right\}$$

$$=\frac{2\tilde{f}_0\lfloor 2\tilde{\tau}_0 K\rfloor}{T}\mathrm{E}_x\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x|^2+1\right)+\frac{K}{T}\log\left(\pi e\sigma_n^2\right)$$

$$=h'_L(\mathbf{y}|\mathbf{x}) \qquad (9.71)$$

where we use the sign $\gtrsim$ to indicate that this lower bound holds only approximately.

9.2.3 The Achievable Rate

Based on the upper and lower bounds on $h'(\mathbf{y})$ and on $h'(\mathbf{y}|\mathbf{x})$ we give upper and lower bounds on the achievable rate with i.d. input symbols in temporal and in subcarrier domain within this section. On the one hand, we give these bounds for the case of a peak power constraint and, on the other hand, we

evaluate the bounds on the achievable rate for the special case of using i.i.d. zero-mean proper Gaussian inputs.

9.2.3.1 Upper Bound

Gaussian Input Distribution

For the special case of i.i.d. zero-mean proper Gaussian input symbols, an upper bound on the achievable rate is given by

$$
\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} = \sup_{\alpha \in [0,1]} \left\{ h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\big|_{\mathrm{PG}} \right\}
$$

$$
= \sup_{\alpha \in [0,1]} \left\{ \frac{K}{T} \log\left(\frac{T}{K}\alpha\rho + 1 \right) \right.
$$

$$
\left. - \frac{2\tilde{f}_0 \lfloor 2\tilde{\tau}_0 K \rfloor}{T} \int_{z=0}^{\infty} \log\left(\frac{T}{4\tilde{f}_0\tilde{\tau}_0 K}\alpha\rho z + 1 \right) e^{-z}dz \right\}
$$

$$
\overset{(a)}{=} \frac{K}{T} \log\left(\frac{T}{K}\rho + 1 \right) - \frac{2\tilde{f}_0 \lfloor 2\tilde{\tau}_0 K \rfloor}{T} \int_{z=0}^{\infty} \log\left(\frac{T}{4\tilde{f}_0\tilde{\tau}_0 K}\rho z + 1 \right) e^{-z}dz
$$

$$
= \mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} \gtrsim \mathcal{I}'(\mathbf{y};\mathbf{x})\big|_{\mathrm{PG}} \tag{9.72}
$$

which holds only approximately due to the fact that we do not have infinitely many subcarriers. Here ρ is the average SNR as defined in (9.37). The supremum is achieved for $\alpha = 1$, i.e., the upper bound is maximized while using the maximum average transmit power. The proof of (a) is analog to the frequency flat case discussed in Section 3.4.1.2. Thus, we have found an approximative upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for the case of a frequency-selective fading channel under the assumption of a brick-shaped scattering function. To the best of our knowledge, this is the only available (approximative) upper bound on the achievable rate with i.i.d. non-peak power constrained input symbols. Other bounds, e.g., the bounds on the capacity given in [33] make explicit use of the peak power constraint in their derivation and therefore cannot be applied to non-peak power constrained input distributions as proper Gaussian inputs. On the other hand, it has to be stated that the bounds given in [33] are bounds on the capacity holding for arbitrary scattering functions of compact support, while our approximative upper bound is only an upper bound on the achievable rate, as we assume i.d. input symbols in temporal and subcarrier domain, which has been required within our derivation. Furthermore, for our derivation, we have to restrict to brick-shaped scattering functions, which is further approximated due to the fact that we do not have infinitely many subcarriers.

As in the flat-fading case, we can modify the approximative upper bound on the achievable rate given in (9.72) based on the fact that the achievable rate is always upper-bounded by the coherent mutual information rate $\mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})$, i.e., cf. (3.87)

$$\mathcal{I}'_{U_{mod}}(\mathbf{y}; \mathbf{x}) = \min\{\mathcal{I}'_U(\mathbf{y}; \mathbf{x}), \mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})\}. \tag{9.73}$$

As in case of perfect CSI at the receiver i.i.d. zero-mean proper Gaussian input symbols are capacity-achieving, we get

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})\big|_{\mathrm{PG}} = \sup_{\mathcal{P}_{\mathrm{OFDM}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) = \frac{K}{T} \int_{z=0}^{\infty} \log\left(1 + \frac{PT}{K}\frac{\sigma_h^2}{\sigma_n^2}z\right) e^{-z}dz$$

$$= \frac{K}{T} \int_{z=0}^{\infty} \log\left(1 + \frac{T}{K}\rho z\right) e^{-z}dz. \tag{9.74}$$

Effectively, this is the capacity of K parallel Rayleigh flat-fading channels, each having an input power PT/K. Furthermore, to get to the information rate, we normalize by T, the duration of one time slot, respectively one OFDM symbol.

Peak Power Constraint

Based on the peak power constrained input set $\mathcal{P}_{\mathrm{OFDM,\ i.d.}}^{\mathrm{peak}}$ in (9.42) we get the following approximative upper bound on the achievable rate with peak power constrained i.d. input symbols

$$\sup_{\mathcal{P}_{\mathrm{OFDM,\ i.d.}}^{\mathrm{peak}}} \mathcal{I}'_U(\mathbf{y}; \mathbf{x}) = \sup_{\mathcal{P}_{\mathrm{OFDM,\ i.d.}}^{\mathrm{peak}}} \{h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\}$$

$$= \sup_{\mathcal{P}_{\mathrm{OFDM,\ i.d.}}^{\mathrm{peak}}} \left\{\frac{K}{T}\log\left(\frac{\alpha\sigma_x^2\sigma_h^2}{\sigma_n^2} + 1\right) - \frac{2\tilde{f}_0\lfloor 2\tilde{\tau}_0 K\rfloor}{T}\mathrm{E}_x\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}|x|^2 + 1\right)\right\}$$

$$\overset{(a)}{\leq} \sup_{\alpha\in[0,1]} \left\{\frac{K}{T}\log\left(\frac{\alpha\sigma_x^2\sigma_h^2}{\sigma_n^2} + 1\right)\right.$$

$$\left. - \frac{2\tilde{f}_0\lfloor 2\tilde{\tau}_0 K\rfloor}{T}\frac{\alpha\sigma_x^2 K}{P_{\mathrm{peak}}}\log\left(\frac{\sigma_h^2}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2}\frac{P_{\mathrm{peak}}}{K} + 1\right)\right\}$$

$$= \sup_{\alpha\in[0,1]} \left\{\frac{K}{T}\log\left(\frac{\alpha PT\sigma_h^2}{K\sigma_n^2} + 1\right) - \frac{2\tilde{f}_0\lfloor 2\tilde{\tau}_0 K\rfloor}{T}\frac{\alpha}{\beta}\log\left(\frac{\sigma_h^2 PT}{4\tilde{f}_0\tilde{\tau}_0\sigma_n^2 K}\beta + 1\right)\right\}$$

$$= \sup_{\alpha\in[0,1]} \left\{\frac{K}{T}\log\left(\frac{T}{K}\alpha\rho + 1\right) - \frac{2\tilde{f}_0\lfloor 2\tilde{\tau}_0 K\rfloor}{T}\frac{\alpha}{\beta}\log\left(\frac{T}{4\tilde{f}_0\tilde{\tau}_0 K}\rho\beta + 1\right)\right\}$$

$$= \mathcal{I}'_U(\mathbf{y}; \mathbf{x})\big|_{P_{\mathrm{peak}}} \tag{9.75}$$

where ρ is the nominal mean SNR as defined in (9.37). The derivation of inequality (a) is based on the same approach as it has been used in (3.70) - (3.74). In addition, we have used (9.36) and the nominal peak-to-average power ratio

$$\beta = \frac{P_{\text{peak}}}{\sigma_x^2 K}.$$ (9.76)

The α maximizing (9.75) is given by

$$\alpha_{\text{opt}} = \min \left\{ 1, \left(\frac{2\tilde{f}_0 \lfloor 2\tilde{\tau}_0 K \rfloor}{K\beta} \log \left(\frac{T}{4\tilde{f}_0 \tilde{\tau}_0 K} \rho\beta + 1 \right) \right)^{-1} - \frac{K}{T\rho} \right\}.$$ (9.77)

Apparently, also this approximative upper bound can be used in combination with the coherent upper bound, cf. (9.73).

9.2.3.2 Lower Bound

No Peak Power Constraint

To get a lower bound on the capacity without a peak power constraint, we have to calculate

$$\sup_{\mathcal{P}_{\text{OFDM, i.d.}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \geq \sup_{\mathcal{P}_{\text{OFDM, i.d.}}} \mathcal{I}'_L(\mathbf{y}; \mathbf{x})$$

$$= \sup_{\mathcal{P}_{\text{OFDM, i.d.}}} \{ h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x}) \}$$

$$\stackrel{(a)}{=} \sup_{\mathcal{P}_{\text{OFDM, i.d.}}} \{ \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - h'_U(\mathbf{y}|\mathbf{x}) \} + h'(\mathbf{y}|\mathbf{h}, \mathbf{x})$$

$$\stackrel{(b)}{=} \sup_{\alpha \in [0,1]} \left\{ \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})\big|_{\text{PG},\alpha} - h'_U(\mathbf{y}|\mathbf{x}) \right\} + h'(\mathbf{y}|\mathbf{h}, \mathbf{x})$$
(9.78)

where for (a) we have used (9.43). Furthermore, (b) is based on the fact that $h'_U(\mathbf{y}|\mathbf{x})$ given in (9.55) only depends on the input distribution in form of the average transmit power αP and the coherent mutual information rate $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ is maximized for i.i.d. zero-mean proper Gaussian inputs. The index PG, α denotes zero-mean proper Gaussian input symbols and an average power equal to αP. Thus, the first term on the RHS of (9.78) corresponds to (9.74) but with an average transmit power of αP.

Thus, with (9.74), (9.44), and (9.55) we get the following lower bound on the channel capacity

$$\sup_{\mathcal{P}_{\mathrm{OFDM,\,i.d.}}} \mathcal{I}'_L(\mathbf{y}; \mathbf{x})$$

$$= \sup_{\alpha \in [0,1]} \left\{ \frac{K}{T} \int_{z=0}^{\infty} \log\left(1 + \frac{T}{K}\alpha\rho z\right) e^{-z} dz - 2f_0 \sum_{i=0}^{K-1} \log\left(1 + \frac{\alpha P}{\sigma_n^2 2 f_0 K} d_i\right) \right\}$$

$$\overset{(a)}{\geq} \frac{K}{T} \int_{z=0}^{\infty} \log\left(1 + \frac{T}{K}\rho z\right) e^{-z} dz - 2\frac{\tilde{f}_0}{T} \sum_{i=0}^{K-1} \log\left(1 + \frac{\rho T}{\sigma_h^2 2 \tilde{f}_0 K} d_i\right) \tag{9.79}$$

$$\overset{(b)}{=} \mathcal{I}'_L(\mathbf{y}; \mathbf{x})\big|_{\mathrm{PG}} \tag{9.80}$$

where for (a) we choose $\alpha = 1$, i.e., use the maximum average transmit power, which does not maximize the argument of the supremum in general. Furthermore, for (a) we have substituted f_0 by its normalized pendant $\tilde{f}_0$, see (9.29). Obviously, the lower bound in (9.79) is achievable with i.i.d. zero-mean proper Gaussian input symbols in temporal and in frequency domain and, thus, (b) holds.

If we use the assumption on a brick-shaped scattering function as it has been used in the derivation of the lower bound on $h'(\mathbf{y}|\mathbf{x})$ in Section 9.2.2.2, the d_i in (9.79) which are defined by (9.54) can be approximated by

$$d_i \approx \begin{cases} \frac{\sigma_h^2}{2\tilde{\tau}_0} & \text{for } 1 \leq i \leq \lfloor 2\tilde{\tau}_0 K \rfloor \\ 0 & \text{otherwise} \end{cases}. \tag{9.81}$$

Based on the assumption of a bricked-shaped scattering function and the approximation in (9.81), we get for (9.79)

$$\sup_{\mathcal{P}_{\mathrm{OFDM,\,i.d.}}} \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) = \frac{K}{T} \int_{z=0}^{\infty} \log\left(1 + \frac{T}{K}\rho z\right) e^{-z} dz$$

$$- 2\frac{\tilde{f}_0}{T} \lfloor 2\tilde{\tau}_0 K \rfloor \log\left(1 + \frac{T}{4\tilde{f}_0\tilde{\tau}_0 K}\rho\right). \tag{9.82}$$

As in the flat-fading case, this lower bound can be modified using the fact that the mutual information rate is nonnegative, cf. (3.94).

Peak Power Constraint

As the lower bound given in (9.79) is based on the coherent capacity which is achieved for zero-mean proper Gaussian inputs, which are non-peak power constrained, we have to modify the lower bound when the input symbols are peak power constrained. Corresponding to the discussion for the flat-fading case in Section 3.4.2.2 a lower bound on the capacity in case of a peak power constraint is given by

$$\sup_{\mathcal{P}^{\text{peak}}_{\text{OFDM, i.d.}}} \mathcal{I}'_L(\mathbf{y};\mathbf{x}) = \sup_{\mathcal{P}^{\text{peak}}_{\text{OFDM, i.d.}}} \{\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) - h'_U(\mathbf{y}|\mathbf{x})\}$$

$$\overset{(a)}{=} \sup_{\mathcal{P}^{\text{peak}}_{\text{OFDM, i.d.}}} \left\{ \frac{1}{T}\mathcal{I}(\tilde{\mathbf{y}};\tilde{\mathbf{x}}|\tilde{\mathbf{h}}) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) - h'_U(\mathbf{y}|\mathbf{x}) \right\}$$

$$\overset{(b)}{\geq} \frac{K}{T}\mathcal{I}(y[n,k];x[n,k]|h[n,k])\big|_{\text{CM},\frac{PT}{K}} - 2\frac{\tilde{f}_0}{T}\sum_{i=0}^{K-1}\log\left(1 + \frac{\rho T}{\sigma_h^2 2\tilde{f}_0 K}d_i\right) \tag{9.83}$$

$$\overset{(c)}{=} \frac{K}{T}\mathcal{I}(y[n,k];x[n,k]|h[n,k])\big|_{\text{CM},\frac{PT}{K}} - 2\frac{\tilde{f}_0}{T}\lfloor 2\tilde{\tau}_0 K\rfloor\log\left(1 + \frac{T}{4\tilde{f}_0\tilde{\tau}_0 K}\rho\right) \tag{9.84}$$

$$= \mathcal{I}'_L(\mathbf{y};\mathbf{x})\big|_{P_{\text{peak}}} \tag{9.85}$$

where (a) results from the fact that the supremum of the coherent mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ is achieved for independent OFDM symbols and for (b) we have used that $\mathcal{I}(\tilde{\mathbf{y}};\tilde{\mathbf{x}}|\tilde{\mathbf{h}})$ corresponds to the coherent mutual information of K parallel flat fading channels. In case of a peak power constraint, we lower-bound it based on the assumption of constant modulus signaling, see Section 3.4.2.2. Thus, $\mathcal{I}(y[n,k];x[n,k]|h[n,k])\big|_{\text{CM},\frac{PT}{K}}$ is the coherent mutual information of a subcarrier when using i.i.d. circularly symmetric constant modulus input symbols with power PT/K. As in the flat-fading case, the coherent mutual information with constant modulus inputs has to be evaluated numerically. In addition, equality (c) in (9.84) holds for the special case of a brick-shaped scattering function corresponding to (9.82). It is based on the approximation in (9.81).

9.2.3.3 The Effect of the Channel Bandwidth B

The bounds on the achievable rate can also be expressed in terms of the bandwidth B. The bandwidth B is given by

$$B = KF \tag{9.86}$$

where K is the number of frequency subcarriers and F is their frequency spacing. For simplicity, from now on we use the following approximation

$$\lfloor 2\tilde{\tau}_0 K\rfloor \approx 2\tilde{\tau}_0 K. \tag{9.87}$$

For a sufficiently large number of subcarriers K, i.e., large bandwidths B, this approximation works well, i.e., it has a small influence on the bounds on the achievable rate.

Based on the bandwidth B, we get the following expressions for the approximative upper bound and the lower bound on the achievable rate with

i.d. input symbols, a peak power constraint, and a brick-shaped scattering function

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{P_{\text{peak}}}$$
$$= \sup_{\alpha \in [0,1]} \left\{ \frac{B}{TF} \log \left(\frac{TF}{B} \alpha \rho + 1 \right) - \frac{4\tilde{f}_0 \tilde{\tau}_0 B}{TF} \frac{\alpha}{\beta} \log \left(\frac{TF}{4\tilde{f}_0 \tilde{\tau}_0 B} \rho \beta + 1 \right) \right\}$$
$$(9.88)$$

$$\mathcal{I}'_L(\mathbf{y};\mathbf{x})\big|_{P_{\text{peak}}}$$
$$= \frac{B}{TF} \mathcal{I}(y[n,k]; x[n,k] | h[n,k]) \big|_{\text{CM}, \frac{PTF}{B}} - \frac{4\tilde{f}_0 \tilde{\tau}_0 B}{TF} \log \left(1 + \frac{TF}{4\tilde{f}_0 \tilde{\tau}_0 B} \rho \right).$$
$$(9.89)$$

For the case of i.i.d. zero-mean proper Gaussian inputs and a brick-shaped scattering function, we get

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\text{PG}}$$
$$= \frac{B}{TF} \log \left(\frac{TF}{B} \rho + 1 \right) - \frac{4\tilde{f}_0 \tilde{\tau}_0 B}{TF} \int_{z=0}^{\infty} \log \left(\frac{TF}{4\tilde{f}_0 \tilde{\tau}_0 B} \rho z + 1 \right) e^{-z} dz$$
$$(9.90)$$

$$\mathcal{I}'_L(\mathbf{y};\mathbf{x})\big|_{\text{PG}}$$
$$= \frac{B}{TF} \int_{z=0}^{\infty} \log \left(1 + \frac{TF}{B} \rho z \right) e^{-z} dz - \frac{4\tilde{f}_0 \tilde{\tau}_0 B}{TF} \log \left(1 + \frac{TF}{4\tilde{f}_0 \tilde{\tau}_0 B} \rho \right).$$
$$(9.91)$$

Here we have to give an important remark. Note that the first term on the RHS of (9.91) corresponds to the coherent capacity given in (9.74). The coherent capacity is achieved for i.i.d. zero-mean proper Gaussian input symbols, i.e., independent symbols in time and in frequency. Obviously, for $B \to \infty$ the first term on the RHS of (9.91) is non-zero. But this is in contradiction to a result given in [78], stating that large bandwidths cannot effectively utilized by spreading the power uniformly in time and frequency. They show that for such input signals the capacity converges to zero for infinite bandwidth, which obviously is in contradiction to our result. We do not know the reason for disagreement. It might be a result of the orthonormal basis that has been chosen for the discretization. In [78] a different discretization has been chosen. Such a discrepancy has also been observed in [33], and there also the different discretization is mentioned as a possible reason. However, the numerical evaluations in Section 9.2.3.4 will show that for small to medium bandwidths the expression for the coherent capacity in (9.74) shows reasonable results.

9.2.3.4 Numerical Evaluation

In this section, we numerically evaluate and compare the approximative upper bound and the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols in temporal and in subcarrier domain given in (9.90) and in (9.91). Furthermore, we compare the approximative upper bound on the achievable rate with i.d. peak power constrained input symbols given in (9.88) with the approximative upper bound for i.i.d. zero-mean proper Gaussian inputs (9.90). Corresponding to the numerical evaluation in [33], we choose the following system parameters which apply for an IEEE 802.11a system with a transmit power of $200\,\mathrm{mW}$, a very strong pathloss of $118\,\mathrm{dB}$, and a receiver noise figure of $5\,\mathrm{dB}$,

- a symbols duration of $T = 0.35\,\mathrm{ms}$ and a subcarrier spacing $F = 3.53\,\mathrm{kHz}$, yielding $TF \approx 1.25$. [3]
- the receive power is normalized by the noise spectral density to

$$\frac{P}{1\,\mathrm{W/Hz}} = 2.42 \cdot 10^7 \frac{1}{\mathrm{s}}. \tag{9.92}$$

Fig. 9.1 shows the approximative upper bound and the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (9.90) and (9.91) for the previously given system parameters and a flat brick-shaped scattering function. Concerning the support of the scattering function, we consider the following two cases:

- A maximum delay $\tau_0 = 0.5\,\mu\mathrm{s}$ and a maximum Doppler spread of $f_0 = 5\,\mathrm{Hz}$, leading to the spread $\Delta_{\mathbb{H}} = 4f_0\tau_0 = 10^{-5}$.
- A large spread of $\Delta_{\mathbb{H}} = 4f_0\tau_0 = 10^{-3}$ to show the effect of the channel dynamics and the delay spread.

Notice, that these bounds are based on the assumption that ICI and ISI are neglected, see Section 9.1.2. For the parameter range evaluated here, i.e., the low SNR, the AWGN will be the dominating noise component and negligence of the interference is reasonable.

In addition, in Fig. 9.1 also the capacity in case of perfect channel state information is shown. It can be clearly observed, that for the case of no channel state information at the receiver side the achievable rate increases with an increasing bandwidth up to a certain bandwidth, which is typically named *critical bandwidth* [33], and then decreases for a further bandwidth extension. For $B \to \infty$ the achievable rate again approaches zero. This behavior can be explained as follows. Up to the critical bandwidth the achievable rate increases due to additional degrees of freedom of the channel that can be used for communication increases. A further extension of the bandwidth leads to a reduction of the achievable rate. This effect occurs due to the increase in

---

[3] This also corresponds to the general guideline on the choice of the ratio $\frac{T}{F} = \frac{\tau_0}{f_0}$ given in [33], [64], [65], and [76].
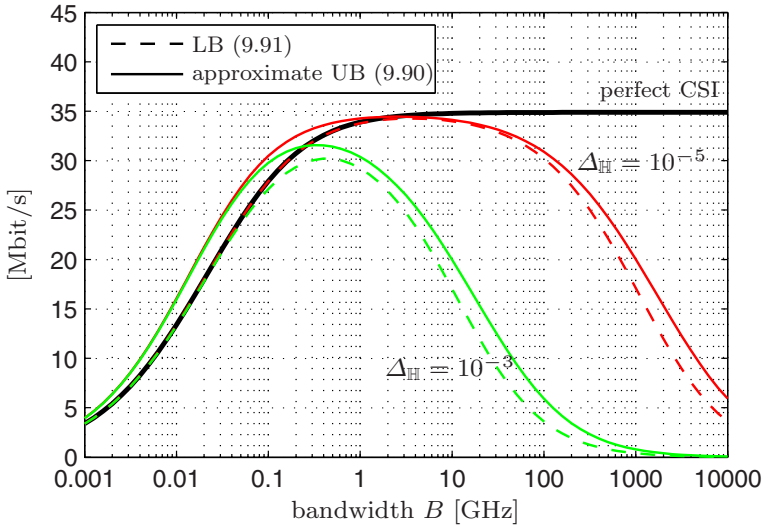
Fig. 9.1: **Bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs** in temporal and in subcarrier domain; perfect CSI corresponds to coherent capacity $\sup_{\mathcal{P}_{\mathrm{OFDM}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ (9.74)

the number of degrees of freedom of the channel, leading, in the context of channel estimation, to an increasing channel estimation error variance. In this regard, consider that the transmitted power is held constant and is transmitted over a larger bandwidth, i.e., the SNR per subcarrier decreases with an increasing bandwidth.

The critical bandwidth obviously decreases with an increasing spread of the scattering function $\Delta_{\mathbb{H}}$. Also the maximum of the achievable rate decreases with an increasing spread $\Delta_{\mathbb{H}}$. This corresponds to the previous explanations. An increasing spread $\Delta_{\mathbb{H}}$ corresponds to a less correlated channel in the temporal and in the frequency domain, i.e., more degrees of freedom of the channel, and, thus, in the context of channel estimation, to a larger channel estimation error variance and a decreased rate.

Furthermore, it has to be considered that the achievable rate is always upper-bounded by the perfect CSI capacity, see (9.73). Up to a specific bandwidth, the lower bound and the perfect CSI capacity virtually coincide. However, there is still a gap, which cannot be identified due to the scaling in Fig. 9.1.

Fig. 9.2 shows the approximative upper bound on the achievable rate with i.d. peak power constrained input symbols (9.88) in comparison to the approximative upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs (9.90), which also holds for i.d. input symbols. As it has al-
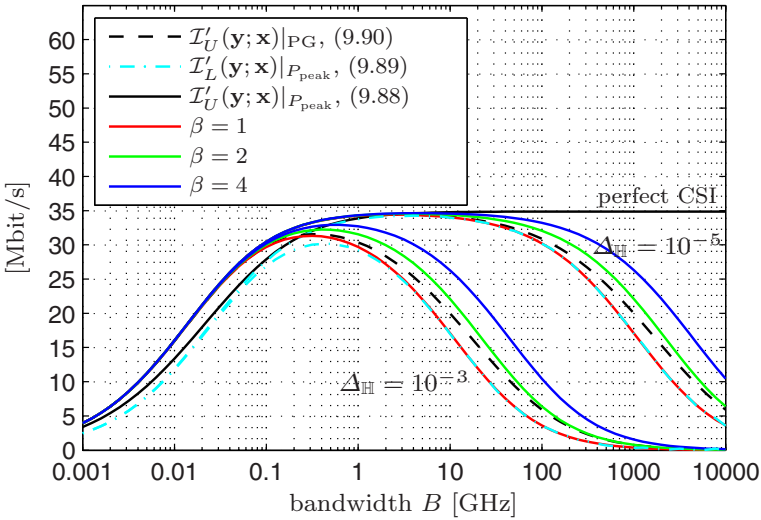
Fig. 9.2: **Comparison of approximative upper bounds on the achievable rate with i.d. inputs for a peak power constraint in (9.88) and with zero-mean proper Gaussian inputs in (9.90)**; notice, (9.90) holds also for i.d. input symbols; nominal peak to average power ratios β; perfect CSI corresponds to coherent capacity $\sup_{\mathcal{P}_{\text{OFDM}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ (9.74); in addition the **lower bound on the peak power constrained capacity in (9.89)** is shown

ready been observed in the flat-fading case in Section 3.4.1, with an increasing nominal peak-to-average power ratio β the approximative upper bound in (9.88) is increasing. For $\beta \to \infty$ the second term in (9.88) approaches zero and thus the bound becomes loose. For $\beta = 2$ the peak power constrained approximative upper bound (9.88) is already larger than the approximative upper bound on the achievable rate for zero-mean proper Gaussian inputs (9.90). Furthermore, the lower bound on the peak power constrained capacity in (9.89) is shown. As this lower bound is by construction also a lower bound on the achievable rate with constant modulus input symbols, it becomes loose towards smaller bandwidths, as in this case the subcarrier SNR increases. With increasing subcarrier SNR the achievable rate with constant modulus input symbols stays increasingly below the capacity.

9.2.3.5 Comparison to Capacity Bounds in [33]

In [33] bounds on the capacity of peak power constrained frequency-selective fading channels are given. Now, we compare these capacity bounds to the bounds on the achievable rate with peak power constrained i.d. input symbols given in (9.83) and (9.88).

Therefore, note that corresponding to the flat-fading case in Section 3.6, the lower bound in (9.83) can be enhanced by using the time sharing argumentation and, thus, by skipping the assumption on i.d. input symbols in temporal domain. This leads to

$$
\sup_{\mathcal{P}^{\mathrm{peak}}_{\mathrm{OFDM}}} \mathcal{I}'_L(\mathbf{y};\mathbf{x})
$$

$$
\geq \max_{\gamma\in[1,\beta]} \left\{ \frac{1}{\gamma T}\mathcal{I}(\tilde{\mathbf{y}};\tilde{\mathbf{x}}|\tilde{\mathbf{h}})\big|_{\mathrm{CM},\gamma PT} - \frac{1}{\gamma}2\frac{\tilde{f}_0}{T}\sum_{i=0}^{K-1}\log\left(1+\frac{\gamma\rho T}{\sigma_h^2 2\tilde{f}_0 K}d_i\right)\right\}
$$
$$(9.93)$$

where $\mathcal{P}^{\mathrm{peak}}_{\mathrm{OFDM}}$ corresponds to $\mathcal{P}^{\mathrm{peak}}_{\mathrm{OFDM,\ i.d.}}$ in (9.42) but without the restriction to i.d. input symbols in temporal domain. The lower bound on capacity in (9.93) exactly corresponds to the lower bound given in [33, (48)].

For the special case of a brick-shaped scattering function also the upper bound on the capacity given in [33, Theorem 1] is equal to the approximative upper bound on the achievable rate with i.d. input symbols given in (9.88), which thus obviously is not only an approximation but a real upper bound. However, the upper bound given in [33] is an upper bound on capacity, as for its derivation no further assumptions, like the assumption on i.d. input symbols as it has been required for the derivation of (9.88), have been made. Furthermore, the upper bound in (9.88) only holds for the special case of a brick-shaped scattering function and holds only approximately as the asymptotic equivalence of Toeplitz and circulant matrices would only hold for infinite bandwidth. In contrast, the capacity upper bound given in [33] has been rigorously proven and holds for arbitrary scattering functions. However, the derivation of the upper bound given in (9.88) based on the lower bound on $h'(\mathbf{y}|\mathbf{x})$ in (9.71) is new, and its derivation is not explicitly based on the assumption of a peak power constraint. Thus, it enables the derivation of an upper bound on the achievable rate with zero-mean proper Gaussian input symbols given in (9.90). In contrast to this, the approach taken to derive the upper bound on capacity given in [33] is inherently based on the peak power constraint, which is used in combination with the relation between mutual information and MMSE given in [45], and cannot be applied to non-peak power constrained input distributions.

9.3 Comparison with Pilot Based Synchronized Detection

In this section, we compare the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input signals in (9.90) and (9.91) to the achievable rate with synchronized detection where the channel estimation is solely based on pilot symbols. Bounds on the achievable rate with synchronized detection using a solely pilot based channel estimation for an OFDM system have been given in [47] and [4]. In the following we recall these bounds to attain a self contained exposition. In this section, we assume that the covariance in frequency domain and in time domain are independent, i.e.,

$$r_{\mathbb{H}}(t,\nu) = r_{\mathbb{H},F}(t)r_{\mathbb{H},D}(\nu) \tag{9.94}$$

where $r_{\mathbb{H},F}(t)$ describes the temporal correlation of the channel fading weights for the individual delay paths, and $r_{\mathbb{H},D}(\nu)$ describes the correlation of the subcarriers in frequency domain. Thus, (9.94) means that all delay paths have an identical temporal correlation. This assumption corresponds to (9.59) as it has been used for the derivation of the lower bound on $h'(\mathbf{y}|\mathbf{x})$.

Based on the temporal autocorrelation function the PSD of the channel fading process of the individual delay tap is given by

$$S_h(f) = \int_{t=-\infty}^{\infty} r_{\mathbb{H},F}(t)e^{-j2\pi tf}\,dt. \tag{9.95}$$

Furthermore, the power delay profile is given by

$$\rho(\tau) = \int_{\nu=-\infty}^{\infty} r_{\mathbb{H},D}(\nu)e^{j2\pi\nu\tau}\,d\nu. \tag{9.96}$$

Due to (9.94) also the following relation holds, cf. (9.7)

$$S_{\mathbb{H}}(f,\tau) = S_h(f)\rho(\tau). \tag{9.97}$$

Corresponding to the flat-fading case, we can rewrite the channel input-output relation using the pilot based channel estimate $\hat{\tilde{\mathbf{H}}}_{\text{pil},n}$ as follows, cf. (9.13)

$$\begin{aligned}
\tilde{\mathbf{y}}_n &\stackrel{(a)}{=} \tilde{\mathbf{H}}_n\tilde{\mathbf{x}}_n + \tilde{\mathbf{n}}_n + \tilde{\mathbf{i}}_n \\
&\stackrel{(b)}{=} \hat{\tilde{\mathbf{H}}}_{\text{pil},n}\tilde{\mathbf{x}}_n + \tilde{\mathbf{E}}_{\text{pil},n}\tilde{\mathbf{x}}_n + \tilde{\mathbf{n}}_n + \tilde{\mathbf{i}}_n
\end{aligned} \tag{9.98}$$

where for (a) we still consider the ICI and ISI terms, i.e., see (9.12)

$$\left[\tilde{\mathbf{i}}_n\right]_k = \sum_{\substack{n'=-\infty \\ (n',k')\neq(n,k)}}^{\infty} \sum_{k'=0}^{K-1} x[n',k']\langle \mathbb{H}g_{n',k'}, g_{n,k}\rangle \tag{9.99}$$

and for (b) we expressed the diagonal channel fading matrix $\tilde{\mathbf{H}}_n$ by its estimate $\hat{\tilde{\mathbf{H}}}_{\mathrm{pil},n}$ and the estimation error $\tilde{\mathbf{E}}_{\mathrm{pil},n}$, cf. (5.9). The variance of the interference on the individual time-frequency slots is given by

$$\sigma_i^2 = \mathrm{E}\left[\left|[\mathbf{i}_n]_k\right|^2\right] \tag{9.100}$$

where it is assumed to be independent of the specific time-frequency slot.
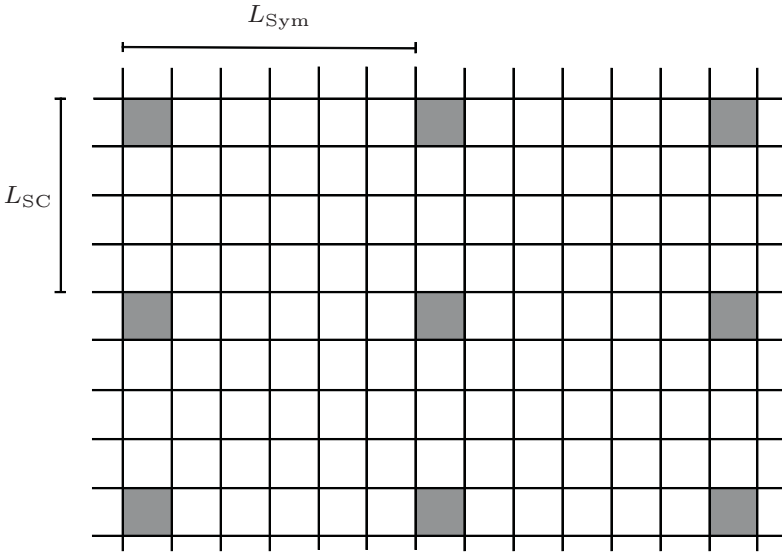
Fig. 9.3: Example of the pilot grid with $L_{\mathrm{Sym}} = 6$ and $L_{\mathrm{SC}} = 4$; pilot symbols in gray, data symbols in white

We assume that dedicated pilot symbols for channel estimation are regularly inserted into the OFDM symbols. Each L_{SC}-th subcarrier is used for a pilot symbol in each L_{Sym}-th OFDM symbol, see Fig. 9.3. Furthermore, we presume that the number of OFDM symbols N is a multiple of L_{Sym} and that the number of subcarriers K is a multiple of the pilot subcarrier spacing L_{SC}. Then, the transmitted block of N OFDM symbols contains

$$NK\left(1 - \frac{1}{L_{\mathrm{Sym}}L_{\mathrm{SC}}}\right) \tag{9.101}$$

data symbols and $NK/(L_{\mathrm{Sym}}L_{\mathrm{SC}})$ pilot symbols.

Recall that the scattering function $S_{\mathbb{H}}(f,\tau)$ has a support that is limited to $[-f_0, f_0] \times [-\tau_0, \tau_0]$, see (9.8). As it has been discussed for the flat-fading case, the pilot spacing has to be chosen such that the channel sampling fulfills the Nyquist criterion to achieve a sufficient channel estimation quality. Concerning the temporal domain, therefore the pilot spacing L_{Sym} must satisfy

$$L_{\text{Sym}} \leq \frac{1}{2f_0 T} = \frac{1}{2\tilde{f}_0}. \tag{9.102}$$

Also concerning the sampling on the frequency domain the Nyquist criterion has to be fulfilled. As we assume that the delay spread is limited to $[-\tau_0, \tau_0]$, each L_{SC}-th subcarrier has to carry pilot symbols with

$$L_{\text{SC}} \leq \frac{1}{2\tau_0 F} = \frac{1}{2\tilde{\tau}_0}. \tag{9.103}$$

The pilot symbols are assumed to have constant modulus, each having a power of σ_p^2.

9.3.1 Channel Estimation

Now, we shortly describe the principle of LMMSE channel estimation in the context of an OFDM system and briefly recall the derivation of an expression for the channel estimation error variance given in [4] and [47]. The channel estimation is solely based on pilot symbols. Therefore, we introduce the vectors $\mathbf{x}_P$ and $\mathbf{y}_P$ corresponding to $\mathbf{x}$ and $\mathbf{y}$ given in (9.20) and (9.21), but only containing the time-frequency slots used for pilot symbols.

In principle, LMMSE channel estimation in an OFDM system is done in three steps [4]. In the first step by multiplication of the elements of received pilot symbol vector $\mathbf{y}_P$ with the corresponding elements of $\mathbf{x}_P$, the modulation is removed in the frequency domain. In the second step an IDFT is performed, giving snapshots of the time domain channel impulse response. These snapshots are then interpolated in the time domain by Wiener filtering. In the third step the estimated channel impulse responses are interpolated in the frequency domain by a DFT transformation of the zero padded impulse responses.

In the following, we assume a rectangular PSD of the channel fading process, i.e.,

$$S_h(\tilde{f}) = \begin{cases} \frac{1}{2\tilde{f}_0} & \text{for } |\tilde{f}| \leq \tilde{f}_0 \\ 0 & \text{otherwise} \end{cases} \tag{9.104}$$

which is normalized to the OFDM symbol duration, see (9.29).

In the discrete-time signal representation the channel consists of

$$\kappa = \lceil B2\tau_0 \rceil \tag{9.105}$$

channel delay taps. The samples of the power delay profile are given by ρ_l, $l = 1, \ldots, \kappa$, cf. (9.96), with

$$\sum_{l=1}^{\kappa} \rho_l = \sigma_h^2. \tag{9.106}$$

We assume a uniform power delay profile, i.e., $\rho_l = \sigma_h^2/\kappa$. This assumption in combination with the rectangular PSD $S_h(\tilde{f})$ in (9.104) approximates the case of a brick-shaped scattering function given in (9.56).

Under these assumptions and for an infinitely long observation interval, i.e., infinitely many OFDM symbols, the estimation error variance for an individual time-frequency slot is given by [4], [47]

$$\sigma_{e_{\text{pil}}}^2 = \mathrm{E}\left[\left|h[n,k] - \hat{h}_{\text{pil}}[n,k]\right|^2\right] = \frac{2\tilde{f}_0 L_{\text{Sym}} \sigma_h^2}{2\tilde{f}_0 L_{\text{Sym}} + \frac{K}{\kappa L_{\text{SC}}} \frac{\sigma_h^2 \sigma_p^2}{(\sigma_n^2 + \sigma_i^2)}} \tag{9.107}$$

where σ_p^2 is the power of the pilot symbols.

From now on, we use the following approximation expressing the κ- delay taps in discrete-time based on the continuous time two-sided delay spread $2\tau_0$, cf. (9.105)

$$\kappa = \lceil B2\tau_0 \rceil \approx B2\tau_0 \tag{9.108}$$

with the system bandwidth B. Using the normalized maximum delay $\tilde{\tau}_0$ as defined in (9.30) and the relation (9.86), we can approximate κ by

$$\kappa = K2\tilde{\tau}_0. \tag{9.109}$$

Therefore, the estimation error variance in (9.107) becomes

$$\begin{aligned}
\sigma_{e_{\text{pil}}}^2 &= \frac{2\tilde{f}_0 L_{\text{Sym}} \sigma_h^2}{2\tilde{f}_0 L_{\text{Sym}} + \frac{1}{2\tilde{\tau}_0 L_{\text{SC}}} \frac{\sigma_h^2 \sigma_p^2}{\sigma_n^2 + \sigma_i^2}} \\
&= \sigma_h^2 \frac{4\tilde{f}_0 \tilde{\tau}_0 L_{\text{Sym}} L_{\text{SC}}}{4\tilde{f}_0 \tilde{\tau}_0 L_{\text{Sym}} L_{\text{SC}} + \frac{\sigma_h^2 \sigma_p^2}{\sigma_n^2 + \sigma_i^2}}
\end{aligned} \tag{9.110}$$

where $\frac{\sigma_h^2 \sigma_p^2}{\sigma_n^2 + \sigma_i^2}$ corresponds to the average SNR of the pilot symbols, including interference. Due to Nyquist sampling, this estimation error variance holds for all time-frequency slots.

9.3.2 Achievable Rate with Pilot Based Synchronized Detection

In this section, we recall a lower bound on the achievable rate with solely pilot based synchronized detection given in [4] and [47]. In [4] also an upper bound on the achievable rate with synchronized detection is given. It has already been observed in the flat-fading case that the gap between the upper and the lower bound is relatively small. Therefore, here we discuss only the lower bound.

Following the same argumentation as in Section 5.2, any correlation of the channel estimation error process in temporal and in frequency domain is not exploited while using solely pilot based synchronized detection. Furthermore, this holds also for correlations of the interference term $\mathbf{i}_n$ in (9.98). As the channel estimation error variance in (9.110) is equal for all time-frequency slots and as each subcarrier corresponds to a flat-fading channel, a lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian data symbols is given by, see [4] and cf. (5.37)

$$
\begin{aligned}
\mathcal{I}'\left(\mathbf{x}; \mathbf{y} \middle| \check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) &= \lim_{N \to \infty} \frac{1}{NT} \mathcal{I}\left(\mathbf{x}; \mathbf{y} \middle| \check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) \\
&\geq \frac{K}{T}\left(1 - \frac{1}{L_{\mathrm{Sym}} L_{\mathrm{SC}}}\right) \mathrm{E}_{\hat{h}_{\mathrm{pil}}[n,k]} \log\left(1 + \frac{|\hat{h}_{\mathrm{pil}}[n,k]|^2 \sigma_d^2}{\sigma_{e_{\mathrm{pil}}}^2 \sigma_d^2 + \sigma_n^2 + \sigma_i^2}\right) \\
&= \frac{K}{T}\left(1 - \frac{1}{L_{\mathrm{Sym}} L_{\mathrm{SC}}}\right) \int_{z=0}^{\infty} \log\left(1 + \rho \eta z\right) e^{-z} dz.
\end{aligned}
\tag{9.111}
$$

where the notation of $\check{\hat{\mathbf{h}}}_{\mathrm{pil}}$ corresponds to (5.33). Moreover, σ_d^2 is the average power of the data symbols on the individual subcarriers. Furthermore, it is used that the channel estimates are zero-mean proper Gaussian with variance $\sigma_h^2 - \sigma_{e_{\mathrm{pil}}}^2$. In addition, the last equality in (9.111) follows from the definition of the average SNR ρ in (9.37) and the SNR degradation factor η, which describes the SNR degradation with respect to the case of perfect channel state information at the receiver side, and which is defined by

$$
\eta = \eta_I \frac{(\sigma_h^2 - \sigma_{e_{\mathrm{pil}}}^2)\sigma_d^2}{P\sigma_h^2 \left(1 + \eta_I \frac{\sigma_{e_{\mathrm{pil}}}^2 \sigma_d^2}{\sigma_n^2}\right)}.
\tag{9.112}
$$

The factor η_I in (9.112) describes the SNR degradation factor due to ICI and ISI and is given by

$$
\eta_I = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_i^2}.
\tag{9.113}
$$

Although the following steps are quite similar to the flat-fading case discussed in Section 5.2 and although the bound is only recalled from [4], in the following we give the lower bound on the achievable rate for the case with and without pilot power optimization depending on the bandwidth B. This enables a direct comparison to the achievable rate with i.i.d. input symbols.

9.3.2.1 Equal Pilot and Average Data Symbol Power

In case the average power of the data symbols and the pilot power are chosen to be equal, i.e., the case without pilot-to-data power ratio optimization, the pilot and the average data power are given by

$$\sigma_p^2 = \sigma_d^2 = \sigma_x^2. \tag{9.114}$$

With (9.36), (9.37), (9.86), (9.110), (9.112), (9.113), and (9.114) the lower bound on the achievable rate with synchronized detection based on a solely pilot based channel estimation in (9.111) can be expressed as a function of the channel bandwidth B by

$$\mathcal{I}'_L \left(\mathbf{x}; \mathbf{y} \big| \check{\mathbf{h}}_{\mathrm{pil}} \right) \bigg|_{\mathrm{SD}}$$

$$= \frac{B}{TF} \left(1 - \frac{1}{L_{\mathrm{Sym}} L_{\mathrm{SC}}} \right) \int_0^\infty \log \left(1 + z \frac{\eta_I \rho \frac{TF}{B}}{4 \tilde{f}_0 \tilde{\tau}_0 L_{\mathrm{Sym}} L_{\mathrm{SC}} \left(1 + \frac{B}{\eta_I \rho TF} \right) + 1} \right) e^{-z} dz. \tag{9.115}$$

Obviously, the lower bound in (9.115) only depends on the product $L_{\mathrm{Sym}} L_{\mathrm{SC}}$ and not on the individual factors. For the numerical evaluation in Section 9.3.2.4, we choose the $L_{\mathrm{Sym}} L_{\mathrm{SC}}$ such that (9.115) is maximized.

9.3.2.2 Pilot-to-Data Power Ratio Optimization

Corresponding to the flat-fading case the pilot-to-average data power ratio can be optimized, cf. Section 5.2.2 and [4]. The achievable rate depends on the ratio between the pilot and the average data symbol power

$$\nu = \frac{\sigma_p^2}{\sigma_d^2}. \tag{9.116}$$

Following the same argumentation as in Section 5.2.2, it is beneficial to choose the maximum possible pilot spacings in temporal and in frequency domain which fulfill the Nyquist conditions in (9.102) and (9.103) and then to optimize ν.

Obviously, the pilot spacings L_{Sym} and L_{SC} can only be integer numbers. To simplify the analysis, we use the following approximations

$$L_{\mathrm{Sym}} = \frac{1}{2\tilde{f}_0} \tag{9.117}$$

$$L_{\mathrm{SC}} = \frac{1}{2\tilde{\tau}_0}. \tag{9.118}$$

Choosing the pilot spacing as in (9.117) and (9.118), and using (9.116), the SNR degradation factor η becomes

$$\eta = \frac{\eta_I^2 \rho \frac{T^2}{K^2} \nu}{\left(4\tilde{f}_0\tilde{\tau}_0(\nu-1)+1\right)\left[\eta_I\rho\frac{T}{K}(1+\nu)-4\tilde{f}_0\tilde{\tau}_0(1-\nu)+1\right]}. \tag{9.119}$$

The factor η is maximized for the following pilot power to average data power ratio [4]

$$\nu_{opt} = \sqrt{\frac{\left(1-4\tilde{f}_0\tilde{\tau}_0\right)^2 + \eta_I\rho\frac{T}{K}\left(1-4\tilde{f}_0\tilde{\tau}_0\right)}{4\tilde{f}_0\tilde{\tau}_0\left(4\tilde{f}_0\tilde{\tau}_0 + \eta_I\rho\frac{T}{K}\right)}}. \tag{9.120}$$

Notice, (9.119) as well as (9.120) correspond to (5.47) and (5.48) in the flat-fading case.

Finally, for the pilot spacings given in (9.117) and (9.118) the lower bound on the achievable rate in (9.111) becomes

$$\mathcal{I}'\left(\mathbf{x};\mathbf{y}\big|\check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right) \geq B\left(\frac{1}{TF} - 4f_0\tau_0\right)\int_{z=0}^{\infty}\log\left(1+\rho\eta z\right)e^{-z}dz$$

$$= \mathcal{I}'_L\left(\mathbf{x};\mathbf{y}\big|\check{\hat{\mathbf{h}}}_{\mathrm{pil}}\right)\bigg|_{\mathrm{SD},\nu_{opt}} \tag{9.121}$$

where we have used (9.29), (9.30), and (9.86), and where η is given by (9.119) with (9.120).

9.3.2.3 The Interference Power

For the evaluation of the lower bound on the achievable rate with pilot based synchronized detection a statement on the SNR degradation due to interference, i.e., the factor η_I is required. In [4] the SNR degradation factor η_I has been discussed in detail with respect to intercarrier interference (ICI). As the motivation of the discussion of the achievable rate with pilot based synchronized detection is mainly given by the comparison to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given

in (9.90) and (9.91), for which the ISI and ICI terms have been completely neglected, see (9.11) and (9.12), in the following we also ignore the interference term, and thus assume $\eta_I = 1$. A detailed discussion of the ISI and ICI terms is beyond the scope of the present work. For a discussion on the validity of this assumption, we refer to [33]. Recently in [32] the capacity bounds given in [33] have been extended to consider interference. For the case of synchronized detection and a study of the term η_I due to ICI, we refer to [4].

9.3.2.4 Comparison to Achievable Rate with i.i.d. Gaussian Input

In this section, we compare the lower bound on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols in (9.90) and (9.91) for the case of a brick-shaped scattering function. We choose the same system parameters that have been used in Section 9.2.3.4. Notice that for the derivation of the synchronized detection based bound on the achievable rate, the brick-shaped scattering function has been approximated in the delay domain with a discrete-time uniform power delay profile.

For the comparison, it has also to be considered that the bounds on the achievable rate with i.i.d. input symbols given in (9.90) and (9.91) are calculated under the assumption that ICI and ISI can be neglected, see (9.11) and (9.12). Therefore, in the following we assume $\eta_I = 1$.

Fig. 9.4 shows the comparison of the lower bound on the achievable rate with solely pilot based synchronized detection with and without pilot-to-data power ratio optimization to the bounds on the achievable rate with i.i.d. input symbols. For both cases the data symbols are assumed to be zero-mean proper Gaussian distributed.

In Fig. 9.4 it can be observed that the difference between the lower bound on the achievable rate without and with pilot-to-data power ratio optimization is hardly visible, in fact the curves are overlapping. Obviously, with the optimization of the pilot spacings L_{Sym} and L_{SC} we can achieve nearly the same effect as with optimizing the pilot-to-average data power ratio and choosing the maximum possible pilot spacing with respect to the Nyquist sampling of the channel fading process.

Concerning the comparison of the achievable rate with solely pilot based synchronized detection and i.i.d. zero-mean proper Gaussian data symbols to the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, we observe that the first one stays significantly below the second one. The gap increases with the channel spread $\Delta_{\mathbb{H}}$, at least for bandwidths up to the critical bandwidth. With decreasing channel bandwidths, i.e., less degrees of freedom of the channel resulting in smaller channel estimation errors, the bounds on the achievable rate with synchronized detection and with i.i.d. input symbols come closer. In contrast, the gap increases with an in-
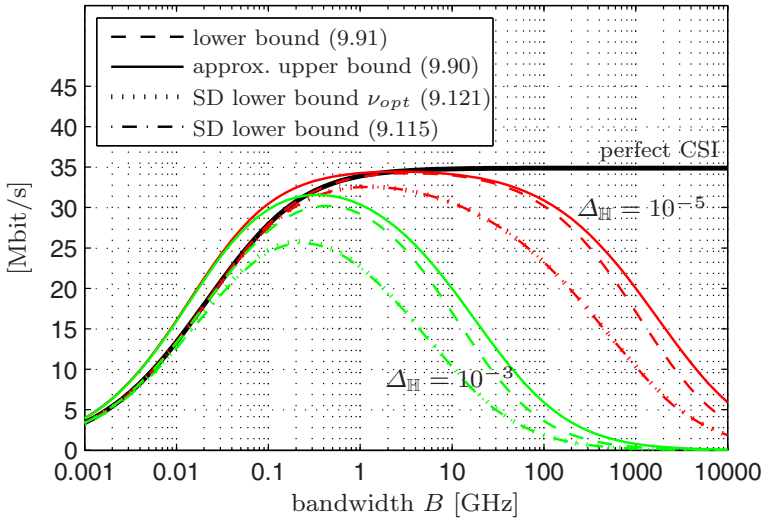
Fig. 9.4: **Comparison of the lower bounds on the achievable rate with synchronized detection (SD) and a solely pilot based channel estimation to the lower bound and the approximative upper bound on the achievable rate with i.i.d. input symbols**; both cases with i.i.d. zero-mean proper Gaussian (data) symbols; for SD bounds without and with pilot power optimization are shown; notice, the lower bounds on the achievable rate with SD with and without optimization of the pilot power are overlapping; assumption $\eta_I = 1$; perfect CSI corresponds to coherent capacity $\sup_{\mathcal{P}_{\mathrm{OFDM}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ (9.74)

creasing system bandwidth B, at least up to reasonable bandwidths, i.e., the critical bandwidth. For $B \to \infty$ all bounds decay to zero. For the design of communication systems this means that for systems using bandwidths corresponding to the critical bandwidth, a significant gain can be achieved while not restricting to a solely pilot based channel estimation for synchronized detection. However, it has to be stated that the majority of communication systems use bandwidths that are significantly smaller than the critical bandwidth (except of ultra-wideband (UWB) systems). In this case, depending on the channel dynamics characterized by $\Delta_{\mathbb{H}}$, the achievable rate with solely pilot based synchronized detection is closer to the achievable rate with i.i.d. input symbols. However, there is still a gap, which due to the scaling in Fig. 9.4 is not visible.

9.4 Summary

In this chapter, we have extended the bounds on the achievable rate with i.d. input symbols given in Chapter 3 for the flat-fading case to the frequency-selective case. Therefore, we have recalled the discrete-time discrete-frequency approximation of the wide-sense stationary uncorrelated scattering channel that has been given in [33]. This approximation works well for highly underspread channels. The main novelty within the present chapter lies in the fact that for the derivation of the upper bound on the achievable rate no use of a peak power constraint has been made, enabling to give an approximate upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. However, due to the finite bandwidth, this upper bound holds only approximately. Beside the evaluation of the bounds for proper Gaussian inputs, we also evaluate the bounds for peak power constrained inputs. The resulting approximate upper bound for i.d. inputs is exactly equivalent to the upper bound on capacity given in [33]. But in contrast to the upper bound on the peak power constrained capacity in [33] the approximate upper bound given in the present work is restricted to brick-shaped scattering functions and is, due to the assumption on i.d. input symbols, not an upper bound on the capacity. The lower bound on capacity with i.d. input symbols can easily transferred to the lower bound on capacity given in [33] by using the time sharing argument and, thus, discarding the assumption on i.d. inputs in temporal domain.

Based on system parameters, corresponding to realistic systems like IEEE 802.11a, we have evaluated the bounds on the achievable rate. In the second part of the present chapter, we have recalled known bounds on the achievable rate with synchronized detection in combination with a solely pilot based channel estimation and compare them to the previously derived bounds on the achievable rate with i.i.d input symbols, assuming in both cases zero-mean proper Gaussian data symbols. This comparison shows that depending on the channel bandwidth and the channel dynamics, the achievable rate with a solely pilot based synchronized detection stays below the achievable rate with i.i.d. input symbols. However, for typical system bandwidths the gap is relatively small in comparison to the achievable rate. If on the other hand the system is operated close to the critical bandwidth, the gap becomes significant. For these comparisons it has to be kept in mind that ISI and ICI has been neglected. Concerning future work, including the interference into the derivations is highly relevant. The interference will have a strong influence on the high SNR behavior, as in this case the interference becomes the dominating factor in comparison to the additive noise. Therefore, we also did not discuss any high SNR behavior within this chapter, as the results would be misleading when ICI and ISI are neglected. Furthermore, in [63] it has been observed that the high SNR behavior of multipath fading channels is very sensitive to the details of the power delay profile of the channel.

Chapter 10
Optimum Discrete Signaling

In this chapter, we consider a time-selective block Rayleigh flat-fading channel. Within a block its temporal correlation corresponds to the stationary fading channel as it has been introduced in Chapter 2. The difference is that after a block of N transmit symbols, the fading state switches to a new independent realization. The stationary fading introduced in Chapter 2 corresponds to the special case of an infinitely large block size N. The motivation to study now a time-selective block Rayleigh flat-fading channel is on the one-hand given by the fact that in many real systems communication on a specific channel resource is of finite length, consider therefore, e.g., a system using frequency hopping. On the other hand, this study enables statements on the achievable rate with receivers that operate blockwise, i.e., where channel estimation as well as detection/decoding is performed blockwise. This also means that for channel estimation within an individual block, no observations of the adjacent blocks are used. As before, we assume that the channel state information is unknown to transmitter and receiver, while the receiver is aware of the channel law. As already stated, even though there exist bounds on the capacity for flat-fading channels, neither exact expressions for the capacity, nor the capacity-achieving input distribution are known. In Chapter 3 and Chapter 4 different bounds on the achievable rate under the assumption of i.d. or i.i.d. input distributions have been derived. These bounds are given on the one hand for a peak power constrained input and on the other hand for zero-mean proper Gaussian input symbols. Thus, the derivation of these bounds is taken over the set of continuous input distributions. In contrast to this, in the present chapter we restrict to discrete input distributions. Furthermore, we drop any assumption on i.d. or i.i.d. input symbols.

Our motivation to study discrete input distributions is twofold. First, in [70], it has been shown that Gaussian input distributions, which are capacity-achieving in case the receiver is aware of the channel state, are in general not capacity-achieving in case the channel state is unknown to the receiver. In contrast, discrete input distributions are capacity-achieving for a variety of conditionally Gaussian channels with bounded input constraints, i.e., with

a peak power constraint, [10]. E.g., for the case of a Rayleigh flat-fading channel without temporal correlation, it has been shown that the capacity-achieving input distribution is discrete with a finite number of mass points [1]. The scenario in the present paper falls into the class of conditionally Gaussian channels. These observations and secondly the fact that practical realizable systems use discrete input distributions are the motivation to study the achievable rates for the given scenario with the restriction to discrete input distributions.

This leads us to the following question: We have a bounded and closed subset $\mathcal{S} \subset \mathbb{C}^N$, where N is the fading block length which corresponds to the length of the transmit sequence in symbols, and a maximum number M of support points $\mathbf{x}_i \in \mathcal{S}$, $i = 1, \ldots, M$, corresponding to the signaling sequences. What is then the optimum choice of the support points and what is their optimum distribution $\mathbf{p} = [p_1, \ldots, p_M]$, with p_i being the probability of transmitting the sequence $\mathbf{x}_i$, that maximizes the mutual information between channel input and output?

In the present chapter, we restrict to the special case where the set $\mathcal{S}$ consists of a fixed amount of predefined support points $\mathbf{x}_1, \ldots, \mathbf{x}_M$ representing the possible transmit sequences. Then, the input distribution $\mathbf{p}$ that maximizes the mutual information can be evaluated. We refer to this input distribution as the *optimum* input distribution. Furthermore, we name the maximum mutual information constrained to a given set of support points *constrained capacity*. In contrast to previous chapters, we use here the term *constrained capacity* instead of the term *achievable rate*, as we explicitly optimize over the input distribution for a given signaling set.

For additive noise channels, this problem has been examined in [35]. We here extend the results in [35] to time-selective block Rayleigh flat-fading channels, where the receiver has no knowledge of the channel state. The channel fading process within a block is characterized by a compactly supported power spectral density (PSD) with a normalized maximum Doppler frequency $f_d < 0.5$, i.e., it is assumed to be *nonregular* [24], [67]. This means that within the block the channel has the same characteristics as described in Chapter 2. However, from block to block it changes to an independent new fading state.

We show that the optimum input distribution $\mathbf{p}$ is characterized by a constant Kullback-Leibler distance between the output PDFs conditioned on the individual input sequences and the unconditional output PDF. For PSK signaling, we explicitly characterize the set of optimum input distributions $\mathbf{p}$. The special case of transmitting one pilot symbol per block, i.e., a symbol that is known to the receiver, and i.i.d. data symbols at all other time instances lies within this set and, thus, is optimal. In addition, we show that the asymptotic high SNR constrained capacity is degraded at least by a factor of $\frac{N-1}{N}$ with respect to the case of perfect channel state information at the receiver. The results of the present chapter have already been published in [25].

The capacity of time-selective block fading channels and also optimum discrete input distributions for noncoherent fading channels have also been studied in prior contributions. The time-selective block fading channel studied in the present chapter is a special case of the block-stationary fading channel, which considers also the more general case that the blocks are correlated between each other, see [71] and [11]. The capacity of time-selective block fading channels and block-stationary fading channels has been discussed, e.g., in [71] and [11]. These publications mainly focus on the behavior of capacity with respect to the SNR, the block length, and in [71] the number of antennas. The first one studies the MIMO case and the second one examines a peak power constrained SISO channel. Examinations of optimum discrete input distributions in the context of noncoherent fading channels have, e.g., been presented in [115] for low SNR MIMO block fading channels. Also the work in [50] on unitary space-time modulation can be mentioned in this context, where motivated by information-theoretic considerations a signaling scheme for the noncoherent MIMO block fading channel is proposed, which is not based on explicit pilot symbols. However, in [50] it is not shown that the proposed constellation design is capacity achieving, instead the motivation is based on a high SNR argumentation.

10.1 Capacity of a Discrete Input Time-Selective Block Fading Channel

As already stated, we use the same system model as it has been introduced for the discrete time flat-fading channel in Chapter 2. The input-output relation is given by (2.1)

$$y_k = h_k x_k + n_k \tag{10.1}$$

and in matrix-vector notation by (2.13)

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} = \mathbf{Xh} + \mathbf{n} \tag{10.2}$$

where the vectors contain N symbols. Within such a block of N symbols the fading has the same statistical properties as described in Chapter 2, i.e., corresponding to a compactly supported PSD with a normalized maximum Doppler frequency $f_d < 0.5$.

Due to the restriction to finite size transmit constellations, the input symbol x_k is a discrete random variable with Q support points. Consequently, the input vector $\mathbf{x}$ is a random variable with

$$M = Q^N \tag{10.3}$$

support points $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{C}^N$.

The mutual information of the channel model given in (10.2) can be calculated as

$$\mathcal{I}(\mathbf{y}; \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}). \tag{10.4}$$

In the following, we examine the constrained mutual information rate of the channel given by (10.2). The constraint is given by the discrete input distribution, where the input vector $\mathbf{x}$ is from the finite set $\mathcal{S}$ given by $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$. We are going to determine the optimum probability distribution $\mathbf{p} = [p_1, \ldots, p_M]$ with p_i the probability of transmitting the sequence $\mathbf{x}_i$, which maximizes

$$C = \frac{1}{N} \max_{\mathbf{p} \in \mathcal{C}} \mathcal{I}(\mathbf{y}; \mathbf{x}) \tag{10.5}$$

where the set $\mathcal{C}$ is convex and given by

$$\mathcal{C} = \left\{ \mathbf{p} = [p_1, \ldots, p_M] \,\middle|\, \sum_{i=1}^{M} p_i = 1, p_i \geq 0 \right\}. \tag{10.6}$$

Now, we discuss the meaning of the quantity C defined in (10.5). For infinite block lengths, i.e., $N \to \infty$, the quantity C in (10.5) corresponds to the constrained capacity of the stationary channel model described in Chapter 2 with the constraint given by the fixed discrete signaling set $\mathcal{S}$. For the identification of (10.5) as the constrained capacity, an infinite block length is necessary. This assumption on an infinite block length (codeword length) is required by the coding theorem, stating that for each rate smaller than the capacity C there exists a code for which the probability of an erroneously decoded codeword approaches zero in the limit of infinite codeword lengths. For a detailed introduction to the concept of Shannon capacity and its proof based on joint typicality we refer, e.g., to [19].

We can also restrict to a finite block length N and assume that coding is performed over an infinite amount of these finite length transmission blocks. In this case, the question on the meaning of the quantity C in (10.5) arises. If we consider a stationary fading process, the channel fading process is correlated over adjacent blocks. Obviously, this correlation is not captured while evaluating C in (10.5). Therefore, in case of a finite block length N the quantity C can be interpreted in the following two ways.

First, for the case of a stationary fading channel as introduced in Chapter 2 it can be understood as a constrained capacity where we introduce the further constraint that the receiver performs detection/decoding and also channel estimation blockwise. Any side information of the channel fading process that is delivered by the observations in the adjacent transmission blocks is not exploited, even not for channel estimation.

Alternatively, it is also possible to argue that for finite N the quantity defined in (10.5) corresponds to the constrained capacity of a time-selective

block fading channel, where now the constraint again is only given by the fixed discrete signaling set $\mathcal{S}$. This time-selective block fading channel has the same characteristics inside the block as the channel model introduced in Chapter 2, but fades independently between the individual blocks of length N. However, this model is obviously not any longer stationary.

Subsequently, we will understand the quantity defined in (10.5) as the constrained ergodic capacity of a time-selective block fading channel with a block length N. The constraint arises from the restriction to the finite set $\mathcal{S}$ with M support points, corresponding to the possible set of signaling sequences.

10.1.1 Optimum Discrete Input Distributions

Now, we study the constrained capacity of the time-selective block fading channel with a discrete input distribution defined in (10.5). This corresponds to evaluate the input distribution $\mathbf{p}$ that maximizes (10.5). Therefore, we first give expressions for the entropies $h(\mathbf{y})$ and $h(\mathbf{y}|\mathbf{x})$ in (10.4).

For the calculation of the channel output entropy conditioned on the channel input $h(\mathbf{y}|\mathbf{x})$, the conditional probability density function $p(\mathbf{y}|\mathbf{x})$ is required. As $\mathbf{y}$ conditioned on $\mathbf{x}$ is zero-mean jointly proper Gaussian, we get

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^N \det\left(\mathbf{R}_{y|x}\right)} \exp\left(-\mathbf{y}^H \mathbf{R}_{y|x}^{-1} \mathbf{y}\right) \tag{10.7}$$

with, cf. (3.20)

$$\mathbf{R}_{y|x} = \mathbf{X}\mathbf{R}_h \mathbf{X}^H + \sigma_n^2 \mathbf{I}_N. \tag{10.8}$$

The distribution of the channel output is given by

$$p(\mathbf{y}) = \sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i). \tag{10.9}$$

Hence, $p(\mathbf{y})$ is a mixture of the conditional densities $p(\mathbf{y}|\mathbf{x}_i)$.

The entropies in (10.4) are given by

$$h(\mathbf{y}) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$$

$$= -\int \sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i) \log \left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \tag{10.10}$$

$$h(\mathbf{y}|\mathbf{x}) = -\int\int p(\mathbf{y},\mathbf{x})\log p(\mathbf{y}|\mathbf{x})d\mathbf{x}\,d\mathbf{y}$$

$$= -\int\sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i)\log p(\mathbf{y}|\mathbf{x}_i)d\mathbf{y}. \qquad (10.11)$$

By interchanging summation and integration, we get for the mutual information

$$\mathcal{I}(\mathbf{y};\mathbf{x}) = -\sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log\left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y}$$

$$+\sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log p(\mathbf{y}|\mathbf{x}_i)d\mathbf{y}$$

$$=\sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y}$$

$$=\sum_{i=1}^{M} p_i D\left(p(\mathbf{y}|\mathbf{x}_i)\left\|\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right.\right) \qquad (10.12)$$

where

$$D\left(g\|h\right) = \int g\log\left(\frac{g}{h}\right) \qquad (10.13)$$

is the Kullback-Leibler distance or the relative entropy between the densities g and h, see also [19].

We now seek to find the distribution $\mathbf{p} = [p_1,...,p_M]$ that maximizes (10.12). This will be pursued in the following by calculating the gradient of $f = \mathcal{I}(\mathbf{y};\mathbf{x})$ with respect to $\mathbf{p}$ and using directional derivatives. However, we would like to point out that there also exists an alternative approach as described at the end of the section.

In the following derivation, we closely follow the lines in [35]. Therefore, we calculate the gradient

$$\nabla f = \left(\frac{df}{dp_i}\right)_{i=1,...,M} \qquad (10.14)$$

with the elements given by

$$\frac{df}{dp_i} = \frac{d}{dp_i} \sum_{k=1}^{M} p_k \int p(\mathbf{y}|\mathbf{x}_k) \left[\log p(\mathbf{y}|\mathbf{x}_k) - \log \left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j) \right) \right] d\mathbf{y}$$

$$= \int p(\mathbf{y}|\mathbf{x}_i) \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} - \int p(\mathbf{y}|\mathbf{x}_i) \log \left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j) \right) d\mathbf{y} - 1.$$

$$(10.15)$$

The directional derivative of f at $\hat{\mathbf{p}} = [\hat{p}_1, \ldots, \hat{p}_M]$ in the direction of $\mathbf{p}$ is given by

$$\langle \nabla f(\hat{\mathbf{p}}), \mathbf{p} - \hat{\mathbf{p}} \rangle = \sum_{i=1}^{M} (p_i - \hat{p}_i) \left(c_i - b_i(\hat{\mathbf{p}}) \right) \tag{10.16}$$

where

$$b_i(\hat{\mathbf{p}}) = \int p(\mathbf{y}|\mathbf{x}_i) \log \left(\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j) \right) d\mathbf{y} \tag{10.17}$$

$$c_i = \int p(\mathbf{y}|\mathbf{x}_i) \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y}. \tag{10.18}$$

Due to the concavity of the log function, $b_i(\hat{\mathbf{p}})$ is concave in $\mathcal{C}$ for $i = 1, \ldots, M$.

It can be shown that the mutual information $f = \mathcal{I}(\mathbf{y}; \mathbf{x})$ is concave in $\mathbf{p}$ on the convex set $\mathcal{C}$, see [38, Theorem 4.4.2]. Thus, the maximum of f is attained at $\hat{\mathbf{p}} \in \mathcal{C}$ if and only if the directional derivatives at $\hat{\mathbf{p}}$ in any direction $\mathbf{p} \in \mathcal{C}$ are nonpositive, i.e.,

$$\sum_{i=1}^{M} (p_i - \hat{p}_i)(c_i - b_i(\hat{\mathbf{p}})) \leq 0. \tag{10.19}$$

Hence, $\hat{\mathbf{p}}$ is an optimum point iff

$$\sum_{i=1}^{M} \hat{p}_i(c_i - b_i(\hat{\mathbf{p}})) = \max_{\mathbf{p} \in \mathcal{C}} \sum_{i=1}^{M} p_i(c_i - b_i(\hat{\mathbf{p}}))$$

$$= \max_{i=1,\ldots,M} (c_i - b_i(\hat{\mathbf{p}})). \tag{10.20}$$

Equality is only achieved if and only if $c_i - b_i(\hat{\mathbf{p}})$ equals some constant χ for all i with $\hat{p}_i > 0$. Recognizing that

$$c_i - b_i(\hat{\mathbf{p}}) = D \left(p(\mathbf{y}|\mathbf{x}_i) \middle\| \sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j) \right) \tag{10.21}$$

we state the following proposition.

Proposition 10.1. *Given the signaling vectors* $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{C}^N$ *for the input variable* $\mathbf{x}$, *the distribution* $\hat{\mathbf{p}}$ *is optimum, i.e., achieves the constrained capacity, if and only if*

$$D\left(p(\mathbf{y}|\mathbf{x}_i) \middle\| \sum_{j=1}^M \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) = \chi \tag{10.22}$$

for some $\chi \in \mathbb{R}$ *and all indices* i *with* $\hat{p}_i > 0$.

Thus, for the optimum $\hat{\mathbf{p}}$ the Kullback-Leibler distance $D\left(p(\mathbf{y}|\mathbf{x}_i) \middle\| \sum_{j=1}^M \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right)$ is constant for all i with positive $\hat{p}_i$. With (10.12), the constrained ergodic capacity amounts to

$$C = \frac{1}{N} \max_{\mathbf{p} \in \mathcal{C}} \mathcal{I}(\mathbf{y}; \mathbf{x}) = \frac{1}{N}\chi. \tag{10.23}$$

Alternatively, we can use the analogy of our problem to the problem of finding the capacity-achieving input distribution of the discrete memoryless channel (DMC). Due to the restriction to a finite set of input sequences with finite length, the Rayleigh fading channel is similar to the DMC, except that its output is continuous. For the DMC a characterization of the capacity-achieving distribution is given in [38, Theorem 4.5.1.]. The same methodology, essentially the Karush-Kuhn-Tucker conditions, may be used in our context, yielding Proposition 10.1.

In conclusion that means, that in order to find the optimum input distribution $\hat{\mathbf{p}}$ for a given signaling set $\mathcal{S}$, we have to find a vector $\mathbf{p}$ that leads to an equal Kullback-Leibler distance $D\left(p(\mathbf{y}|\mathbf{x}_i) \middle\| \sum_{j=1}^M \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right)$ for all its elements i. For general input distributions using phase and magnitude components, e.g., QAM signal constellations, we did not find a closed form solution on the constrained capacity-achieving input distribution. Therefore, in the following we restrict to discrete constant modulus input sets, i.e., PSK signaling, which enable a further analytical treatment.

10.2 Constant Modulus Input Distributions

In this section, we give an explicit characterization of the optimum input distribution $\mathbf{p}$ for the special case of constant modulus input distributions, i.e., PSK type signaling with

$$x_k = \sigma_x \exp\left(j 2\pi \frac{i}{Q}\right), \quad i = 0, \ldots, Q - 1 \tag{10.24}$$

where $\sigma_x^2 = |x_k|^2, \forall k$. Here, j is the imaginary unit, i.e., $j = \sqrt{-1}$. Without loss of generality, in the following we assume $\sigma_x^2 = 1$.

10.2.1 Distinguishable Transmit Sequences

For input signals given by (10.24), the probability density function of the output conditioned on the input sequence (10.7) can be simplified to

$$p(\mathbf{y}|\mathbf{x}_i) = \frac{\exp\left(-\mathbf{y}^H \mathbf{X}_i \left(\mathbf{R}_h + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{X}_i^H \mathbf{y}\right)}{\pi^N \det\left(\mathbf{R}_h + \sigma_n^2 \mathbf{I}_N\right)}. \tag{10.25}$$

It can be shown that the density function conditioned on the two input sequences $\mathbf{x}_m$ and $\mathbf{x}_n$ is equal, i.e.,

$$p(\mathbf{y}|\mathbf{x}_n) = p(\mathbf{y}|\mathbf{x}_m) \tag{10.26}$$

if and only if

$$\mathbf{x}_n = \mathbf{x}_m \exp\left(j\phi\right) \text{ for some } \phi \in [0, 2\pi). \tag{10.27}$$

Thus, transmit sequences fulfilling the constraint given in (10.27) cannot be distinguished by the receiver.

We select a subset $\mathcal{S}_0 \subseteq \mathcal{S}$ of maximal cardinality such that the elements of $\mathcal{S}_0$ are pairwise distinguishable, i.e.,

$$p(\mathbf{y}|\mathbf{x}_n) \neq p(\mathbf{y}|\mathbf{x}_m) \text{ for any } \mathbf{x}_n \neq \mathbf{x}_m \in \mathcal{S}_0. \tag{10.28}$$

It is easy to see that $|\mathcal{S}_0| = \frac{M}{Q} = Q^{N-1}$.

10.2.2 Optimum Input Distribution

Based on $\mathcal{S}_0$, the Kullback-Leibler distance in (10.22) can be transformed to

$$D\left(p(\mathbf{y}|\mathbf{x}_i) \middle\| \sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) = \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y}$$

$$= \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j) \sum_{\{k|\mathbf{x}_k = \mathbf{x}_j e^{j\phi}\}} \hat{p}_k}\right) d\mathbf{y} \tag{10.29}$$

where we have used (10.26) and (10.27). Based on Proposition 10.1 we can now give the following proposition.

Proposition 10.2. *The distribution*

$$\sum_{\{k|\mathbf{x}_k=\mathbf{x}_l e^{j\phi}\}} \hat{p}_k = \psi = \frac{Q}{M} = \frac{1}{Q^{N-1}}, \quad \forall l : \mathbf{x}_l \in \mathcal{S}_0 \tag{10.30}$$

is optimum, i.e., it achieves the constrained capacity.

Intuitively, the optimum input distribution corresponds to a uniform distribution over the space of distinguishable transmit sequences.

Proof. We have to show that for the input distribution given in (10.30), the Kullback-Leibler distance (10.29) is independent of the index i. With (10.21) the Kullback-Leibler distance can be expressed by $c_i - b_i(\hat{\mathbf{p}})$, with c_i and $b_i(\hat{\mathbf{p}})$ given in (10.18) and (10.17).

We first show that the term c_i given in (10.18) is independent of the index i for constant modulus input distributions. All signaling sequences $\mathbf{x}_i$ can be generated based on an arbitrarily chosen vector $\mathbf{x}_1 \in \mathcal{S}$ as

$$\mathbf{x}_i = \mathbf{U}_i \mathbf{x}_1 \tag{10.31}$$

where the matrix $\mathbf{U}_i$ is diagonal, orthonormal and, thus, unitary.

The conditional density $p(\mathbf{y}|\mathbf{x}_i)$, see (10.25), obeys the following property

$$p(\mathbf{y}|\mathbf{x}_i) = p(\mathbf{y}|\mathbf{U}_i \mathbf{x}_1) = p(\mathbf{U}_i^H \mathbf{y}|\mathbf{x}_1). \tag{10.32}$$

With (10.32), we get

$$
\begin{aligned}
c_i &= \int p(\mathbf{y}|\mathbf{x}_i) \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} \\
&= \int p(\mathbf{U}_i^H \mathbf{y}|\mathbf{x}_1) \log p(\mathbf{U}_i^H \mathbf{y}|\mathbf{x}_1) d\mathbf{y} \\
&= \int p(\mathbf{y}|\mathbf{x}_1) \log p(\mathbf{y}|\mathbf{x}_1) d\mathbf{y} \tag{10.33}
\end{aligned}
$$

as $\mathbf{U}_i$ is unitary. Thus, c_i is independent of the index i.

For $b_i(\hat{\mathbf{p}})$ we get with (10.29) and (10.30)

$$
\begin{aligned}
b_i(\hat{\mathbf{p}}) &= \int p(\mathbf{y}|\mathbf{x}_i) \log \left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j) \right) d\mathbf{y} \\
&= \int p(\mathbf{U}_i^H \mathbf{y}|\mathbf{x}_1) \log \left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j) \right) d\mathbf{y} \\
&= \int p(\mathbf{y}|\mathbf{x}_1) \log \left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{U}_i^H \mathbf{x}_j) \right) d\mathbf{y}
\end{aligned}
$$

$$\overset{(a)}{=} \int p(\mathbf{y}|\mathbf{x}_1)\log\left(\psi \sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \tag{10.34}$$

where we used (10.32) and for (a) we applied the following relation

$$\sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{U}_i^H\mathbf{x}_j) = \sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j). \tag{10.35}$$

Equation (10.35) holds, as we sum over all elements $\mathbf{x}_j$ of $\mathcal{S}_0$, which is of maximal cardinality. In addition, let us assume that the diagonal projection matrix $\tilde{\mathbf{U}}_k$ differs from the identity matrix only in one entry. Applying this matrix to all elements $\mathbf{x}_j$ of $\mathcal{S}_0$ leads to a new set $\mathcal{S}_0'$ of distinguishable sequences of maximum cardinality. Considering in addition, that an arbitrary matrix $\mathbf{U}_i$ can be constructed by a product of matrices $\tilde{\mathbf{U}}_k$ where in each one only one element differs from 1, i.e.,

$$\mathbf{U}_i = \prod_k \tilde{\mathbf{U}}_k \tag{10.36}$$

$\mathbf{U}_i^H\mathbf{x}_j$ is a projection into a new set $\mathcal{S}_0'$. As $\sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)$ is equal for all possible sets $\mathcal{S}_0$, equation (10.35) holds.

Therefore, $b_i(\hat{\mathbf{p}})$ is independent of the index i for the distribution (10.30). Finally, we have shown that for the distribution in (10.30), the Kullback-Leibler distance $D\left(p(\mathbf{y}|\mathbf{x}_i)\|\sum_{j=1}^M \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right)$ is constant, and, thus, (10.30) is optimum. $\qquad\square$

With (10.22), (10.23), (10.29), and (10.30) the constrained capacity is given by

$$C = \frac{1}{N}\int p(\mathbf{y}|\mathbf{x}_i)\log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\frac{1}{Q^{N-1}}\sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y}$$

$$= \frac{N-1}{N}\log(Q) - \frac{1}{N}\int p(\mathbf{y}|\mathbf{x}_i)\log\left(\frac{\sum_{\mathbf{x}_j\in\mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_i)}\right) d\mathbf{y}$$

$$= \frac{N-1}{N}\log(Q) - \frac{1}{N}\int p(\mathbf{y}|\mathbf{x}_i)\log\left(1 + \frac{\sum_{\mathbf{x}_j\in\mathcal{S}_0\setminus\mathbf{x}_i} p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_i)}\right) d\mathbf{y}. \tag{10.37}$$

10.2.3 Asymptotic SNR Behavior

As we assume the PSD of the channel fading process to be compactly supported and characterized by a maximum normalized Doppler frequency

$f_d < 0.5$, there are eigenvalues of the channel covariance matrix $\mathbf{R}_h$ which are close to zero, if f_d is not close to 0.5 and if the block length N is sufficiently large. Here, the adjective *sufficiently* strongly depends on f_d. Thus, in this case $\mathbf{R}_h$ is close to singular. As in addition, the sequences constituting the set $\mathcal{S}_0$ are distinguishable, numerical evaluation shows that the second term on the RHS of (10.37) is close to zero and hence,

$$\lim_{\rho \to \infty} C \approx \frac{N-1}{N} \log(Q) \tag{10.38}$$

with ρ being the mean SNR defined in (2.18), for N sufficiently large and f_d sufficiently small for a given constellation size Q. Furthermore, it depends on the constellation size if the behavior described in (10.38) can be observed. For smaller constellation sizes Q a behavior corresponding to (10.38) is seen already for smaller block lengths N as in this case the distance between the signaling points in the signaling space is larger.

Considering perfect channel state information, the mutual information achieves $\log(Q)$ if the SNR is so large that the conditional channel output PDF $p(\mathbf{y}|\mathbf{x}, \mathbf{h})$ becomes so *peaky* that erroneous decoding occurs with probability close to zero. Here, in the noncoherent case, besides the additive Gaussian noise we have the additional random component introduced by the fading channel. Therefore, also in case of an infinite high SNR, there is still random. The amount of this random depends on the PSD $S_h(f)$. For the special case of a rectangular PSD or the Jakes' spectrum the PSD $S_h(f)$ is described by the single parameter f_d. That means that for $f_d \to 0$, the conditional PDFs $p(\mathbf{y}|\mathbf{x}_i)$ in (10.37) also become more peaky. On the other hand, the required *peakiness* resulting in a behavior as given in (10.38) depends on the distance of the signal points in the signaling space, which is a result of the constellation size Q and the block length N. Thus, the behavior depicted in (10.38) corresponds to the case where the randomness introduced by the channel is *sufficiently* small, allowing the PDFs $p(\mathbf{y}|\mathbf{x}_i)$ to become sufficiently peaky when the SNR is increased to infinity. This behavior can already be observed for the parameters $N = 6$ and $f_d = 0.2$ and $Q = 2$ used in the numerical evaluation in Section 10.2.5, see Fig. 10.1.

10.2.4 Interpretation

The optimum input distribution given in (10.30) intuitively corresponds to a uniform distribution over the space of distinguishable transmit sequences. One specific solution, being included in the set of optimum input distributions is to use only distinguishable transmit sequences, i.e., sequences taken from one set $\mathcal{S}_0$, thus fulfilling (10.28), i.e., given by the distribution

$$\hat{p}_i = \begin{cases} \frac{Q}{M} & \text{for } \mathbf{x}_i \in \mathcal{S}_0 \\ 0 & \text{for } \mathbf{x}_i \notin \mathcal{S}_0 \end{cases}. \tag{10.39}$$

As the cardinality of a subset $\mathcal{S}_0$ is Q^{N-1}, the constrained capacity is limited to $\frac{N-1}{N}\log(Q)$, independent of f_d, corresponding to (10.38).

In case the set $\mathcal{S}_0$ is constructed such that all used transmit sequences are characterized by having a fixed symbol at a predetermined time instant, this solution corresponds to the use of one pilot symbol and i.i.d. data symbols at all other time instances[1]. This intuitively explains why at least one signaling dimension, i.e. the information transmitted by one symbol, is lost for providing a phase reference for the receiver.

The above result should not be understood in the way that it is not possible to use all the transmit sequences of the set $\mathcal{S}$. In this case, it has to be assured that the information that is mapped to non-distinguishable sequences $\mathbf{x}$ is equivalent, as the differentiation between these sequences is impossible.

Following the argumentation in Section 10.2.3, and using the fact that for $N \to \infty$ the channel correlation matrix $\mathbf{R}_h$ becomes asymptotically singular for $f_d < 0.5$, for an infinite SNR ρ we get

$$\lim_{N\to\infty} \lim_{\rho\to\infty} C = \log(Q) \tag{10.40}$$

which corresponds to the case where the receiver knows the channel fading process.

10.2.5 Numerical Results

Fig. 10.1 shows the result of a numerical evaluation by Monte Carlo simulation of (10.37) for $Q = 2$, i.e., BPSK, and for $Q = 4$, i.e., QPSK. The temporal correlation of the channel fading process within a block is determined by the Jakes' spectrum with maximum Doppler frequency f_d, see Section 2.1. Thus the corresponding covariance matrix is given by (2.17) and the autocorrelation function $r_h(l) = J_0(2\pi f_d l)$, where J_0 is the zeroth-order Bessel function of the first kind. For comparison also the mutual information in case of perfect channel state information (CSI) is shown.

For infinite SNR, i.e., $\rho \to \infty$, the curves in Fig 10.1 show the behavior described in (10.38). In addition, we see that for a given SNR and a given sequence length, i.e., block length, of N the constrained capacity decreases

---

[1] Note that the statement on i.i.d. data symbols is not a contradiction to the use of a codebook with codewords, which obviously means that data symbols are not independent. When speaking about i.i.d. data symbols to calculate the mutual information, the distribution for the generation of the codebook is meant. Of course, reliable communication in the sense of a diminishing probability of an erroneous decoded codeword for infinite codeword lengths requires the use of a code.

with increasing f_d. Furthermore, it should be noted that the fading block length N influences the constrained capacity. The smaller the block length N is, the smaller is the constrained capacity. Notice, this is not shown in Fig. 10.1.

As the numerical evaluation is based on a Monte Carlo simulation, the calculation time increases with the number of distinguishable transmit sequences Q^{N-1}, i.e., exponentially in N. Therefore, we have restricted to $N = 6$ for numerical evaluation.
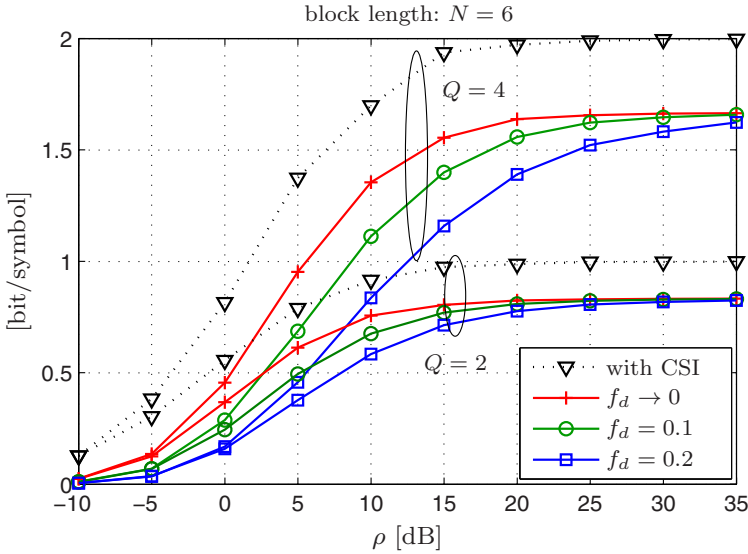
Fig. 10.1: Effect of Q and f_d on the constrained capacity (10.37), numerical evaluation; $S_h(f)$ given by a Jakes' spectrum

10.3 What about Periodic Pilot Symbols in Stationary Fading?

The fact that for a time-selective block fading channel and PSK modulation, it is optimal to use one pilot symbol per block and i.i.d. data symbols at all other time instances allows the following statement. In the context of a stationary fading channel and for the special case of PSK modulation, the use of pilot symbols which are periodically inserted into the transmit sequence, as it has been described in Chapter 5, is not optimal with respect to the achievable rate. This can be seen as the stationary fading channel corresponds

to the case of a time-selective block fading channel with an infinite block length.

However, for practical receiver design also the decoding complexity is a major issue. Therefore, consider that in general the ML-decoding that has to be performed by the receiver to find the transmitted sequence of maximum likelihood means to evaluate

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \{\mathcal{W}_x\}} p(\mathbf{y}|\mathbf{x}) \tag{10.41}$$

where $\mathcal{W}_x$ is the set of all possible codewords. In addition, the conditional output PDF is given by, see (10.7)

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^N \det(\mathbf{R}_{y|x})} \exp\left(-\mathbf{y}^H \mathbf{R}_{y|x}^{-1} \mathbf{y}\right) \tag{10.42}$$

and the correlation matrix $\mathbf{R}_{y|x}$ is given by, cf. (3.20)

$$\mathbf{R}_{y|x} = \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2\mathbf{I}_N. \tag{10.43}$$

The evaluation of (10.41) is obviously quite complex, as, if no approximations are made, $p(\mathbf{y}|\mathbf{x})$ has to be evaluated for all $\mathbf{x}$ contained in $\mathcal{W}_x$.

In contrast, in the coherent case, i.e., if the channel is known, the detection can be simplified as follows

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg \max_{\mathbf{x} \in \{\mathcal{W}_x\}} p(\mathbf{y}|\mathbf{x}, \mathbf{h}) \\
&= \arg \max_{\mathbf{x} \in \{\mathcal{W}_x\}} \frac{1}{\pi^N \sigma_n^{2N}} \exp\left(-\frac{1}{\sigma_n^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right) \\
&= \arg \max_{\mathbf{x} \in \{\mathcal{W}_x\}} \prod_{k=1}^{N} \frac{1}{\pi \sigma_n^2} \exp\left(-\frac{1}{\sigma_n^2} |y_k - h_k x_k|^2\right).
\end{aligned} \tag{10.44}$$

Obviously, here the demapping can be performed per symbol, leading to a much lower decoding complexity.

In case the channel is unknown, it can be estimated based on pilot symbols which are periodically inserted into the transmit symbol sequence. Based on the channel estimates coherent detection can be performed, see Chapter 5. [2] Thus, from an engineering point of view, the use of pilot symbols, which are periodically inserted into the transmit sequence, is a good solution to achieve a high spectral efficiency in combination with receivers of low computational complexity.

Nevertheless, there are various approaches to noncoherent detection and decoding, which, based on different approximations, try to decrease the computational complexity. The use of block based demodulation has been stud-

---

[2] Remark: Due to the channel estimation error, in this case coherent detection corresponds to mismatch decoding, see Section 5.1.2.

ied, e.g., in [23] and [130]. The combination of such a block based demodulation using a noncoherent metric with iterative demodulation and decoding has been studied in a variety of contributions, e.g., to name just a few [18], [73], [72], [51], [86], [85], and [13]. All of these approaches use a serially concatenated scheme, where the inner code corresponds to a modulation coder, which is connected with an outer code by interleaving. On the receiver side demodulation and decoding is then performed iteratively. Some of the works are restricted to constant phase or block fading channels, or at least use a corresponding metric as an approximation in case of a time varying channel, see, e.g., [13], [85], and [18]. Other contributions explicitly consider time varying channels, see, e.g., [72], [51], and [86]. These works solve the problem arising from the exponential complexity of noncoherent detection in various ways. Reductions in receiver complexity can, e.g., be achieved by discretizing the phase space, see, e.g., [13] and [86]. Depending on the choice of the inner code, often referred to as modulation code, the noncoherent demodulation can also be simplified or approximated leading to a reduced decoding complexity. For inner modulation coding, most of the mentioned publications consider a differential encoder.

10.4 Conclusion

In this chapter, we have considered a time-selective block Rayleigh flat-fading channel, where the channel state is unknown to the transmitter and receiver, while the receiver is aware of the channel law. The channel is thus stationary and temporally correlated within a block, whereas between the blocks of length N it fades independently. It has been shown that for a given discrete signaling set, the optimum input distribution is characterized by a constant Kullback-Leibler distance between the output PDFs conditioned on the individual input sequences and the unconditional output PDF. We showed this based on directional derivatives. On the other hand, this fact can also be shown using the analogy of the given scenario to a discrete memoryless channel, for which it is a well known result that the capacity achieving input distribution is characterized by a constant Kullback-Leibler distance. The corresponding proof given in [38] is based on Karush-Kuhn-Tucker conditions. For the special case of PSK signaling sequences, we have derived an explicit expression for the optimum input distribution achieving the constrained capacity, which corresponds to a uniform distribution over the space of distinguishable transmit sequences. Furthermore, we have identified the strategy of transmitting one pilot symbol and i.i.d. data symbols at all other time instances as being included in the set of optimum input distributions. For asymptotic high SNR, the constrained capacity is at least degraded by a factor of $(N-1)/N$ compared to the case of perfect channel state information at receiver side.

The observation that for time-selective block fading the use of a single pilot symbol per block and i.i.d. data symbols at all other time instances is optimal for PSK modulation shows that, at least for PSK modulation, pilot symbols which are periodically inserted into the transmit sequence, as it has been described in Chapter 5, are not optimal with respect to the achievable rate in case of stationary fading. Nevertheless, from an engineering point of view they have the advantage that they allow for coherent detection and, thus, enable receivers with a low computational complexity.

The extension of the approach discussed in the present chapter to signaling sets making use of the amplitude component, e.g., QAM signal constellations would be highly interesting. Furthermore, a general optimization over the input distribution, including the choice of the optimal signaling constellation, is of high relevance.

Chapter 11
Conclusion

The motivation of the present work has been the lack of knowledge on the achievable rate of stationary fading channels, where neither the transmitter nor the receiver has knowledge on the channel fading process, while the receiver is aware of the law of the channel. This channel— which is often referred to as *noncoherent fading channel*— corresponds to the common scenario of a wireless communication system. In many typical communication systems the receiver attains knowledge of the channel fading process based on pilot symbols, i.e., symbols which are known to the receiver and which are regularly introduced into the transmit data stream. These pilot symbols allow for a low complexity channel estimation followed by coherent detection/decoding, often named *synchronized detection*. From an information theoretic point of view, the use of pilot symbols can be understood as a specific type of coding. However, this type of coding has never been proven to be optimal. Therefore, it is highly interesting to know how close the achievable rate of systems based on synchronized detection with a solely pilot based channel estimation is to channel capacity, i.e., the ultimate limit on the data rate still allowing for error-free transmission. While tight[1] bounds on the achievable rate are known in case of using synchronized detection in combination with a solely pilot based channel estimation, the capacity of the corresponding channel, i.e., without any restricting assumptions on pilot symbols or the applied receiver structure, is still a topic of research. Besides the general interest in the capacity of the noncoherent channel, the emerging research on receivers based on iterative synchronization and decoding, i.e., using reliability information on data symbols to iteratively enhance the channel estimation, strongly motivates the study on the capacity of this kind of channel. The reason for this is that the existing bounds on the achievable rate with synchronized detection in combination with solely pilot based channel estimation are not longer valid for iterative code-aided synchronized detection based receivers. Therefore, an

---

[1] The term *tight* means that the bounds are sufficiently close from an engineering point of view and is not meant in the sense of mathematical tightness.

evaluation of the capacity would also allow to study the maximum possible gain while using iterative synchronization and decoding in comparison to solely pilot based synchronized detection. Furthermore, knowledge on the capacity of noncoherent channels is also important concerning other system design parameters like the number of useful antennas in MIMO scenarios, and the dependency of the capacity on the system bandwidth in the context of frequency selective wideband channels.

As the study of the noncoherent capacity turns out to be very challenging, we simplify the problem by introduction of some constraints on the input distribution and, thus, we do not use the term *capacity* but the term *achievable rate*. This means that for the derivation of the bounds on the achievable rate we restrict to i.d. or i.i.d. input symbols. Furthermore, we aim to get bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs, as they are capacity achieving in the coherent scenario. This requires that the derivation of the bounds on the achievable rate does not rely on any peak power constraint. This aspect is different to most of the existing work on capacity bounds for noncoherent fading channels.

In the following, we summarize the specific contributions and findings of the present work. We start with the study of the achievable rate of a discrete-time Rayleigh flat-fading channel with i.d. input symbols, which is characterized by a PSD of the channel fading process that is compactly supported. Furthermore, the channel fading process is assumed to be *nonregular* with a maximum normalized Doppler frequency $f_d < 0.5$. We calculate a new upper bound on the achievable rate with i.d. input symbols for the specific case of a rectangular PSD. The novelty of this bound lies in its derivation which is not based on any assumption of a peak power constraint. Therefore, we can evaluate this bound also for proper Gaussian input symbols. Furthermore, evaluation of this bound for a peak power constraint leads to the same expression, as the upper bound on the peak-power constrained capacity given in [107]. With the upper bound on the achievable rate and a known lower bound on the capacity that we re-derive within this work, we have found a set of bounds, which for i.i.d. zero-mean proper Gaussian inputs is tight in the sense that their difference is bounded for all SNRs. We are able to bound the gap analytically by $(1+2f_d)\gamma$ [nat/cu] with the Euler constant $\gamma \approx 0.577$ and the maximum normalized Doppler frequency f_d. Furthermore, the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols converges to the coherent capacity for asymptotically small channel dynamics and thus becomes tight. In case of a peak power constraint, it is not possible to calculate bounds, which show this behavior, as in this case the lower bound will always be lower than the coherent capacity, also for asymptotically small channel dynamics. This is a result of the fact that the capacity-achieving input distribution in the coherent case is non-peak power constrained. Moreover, the derived bounds show that the achievable rate with i.i.d. zero-mean proper Gaussian input symbols has the same high SNR slope (pre-log) as the peak power constrained channel capacity, which

for a compactly supported PSD is given by $1 - 2f_d$. In addition, it is shown that the peak power constrained capacity upper bound is loose for a nominal peak-to-average power ratio $\beta > 1$ and high SNR, which is not the case for the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. Summing up, for the specific case of zero-mean proper Gaussian inputs we get bounds on the achievable rate, which are tight in the sense of a bounded difference over the whole SNR range. In contrast, other available bounds on capacity mostly are tight only in a specific SNR range, e.g., [107] focuses on the low SNR regime whereas [67] considers the high SNR regime.

For the specific case of i.i.d. input symbols, i.e., with the further restriction to independent input symbols, we derive another upper bound which is based on the one-step channel prediction error variance. The derivation of this bound is based on the fact that the calculation of the channel prediction error variance under the assumption that all past transmit symbols are constant modulus symbols yields an upper bound on the achievable rate. In contrast to other capacity bounds based on this approach, the derivation is not restricted to peak power constrained input symbols. One motivation to study this bound is the link to the physical interpretation of channel prediction. Furthermore, this bound holds for arbitrary compactly supported PSDs of the channel fading process. In contrast, the upper bound on the achievable rate with i.d. input symbols given before holds only for rectangular PSDs. For i.i.d. zero-mean proper Gaussian inputs the comparison of this upper bound on the achievable rate to the one given before shows that their ordering depends on the SNR.

To evaluate the performance of synchronized detection in combination with a solely pilot based channel estimation, we compare its achievable rate to the achievable rate with i.i.d. symbols, i.e., without any assumption on pilot symbols. This comparison shows that for channel dynamics as they are typically observed in mobile environments the achievable rate with a pilot based synchronized detection is lower than the lower bound on the achievable rate with i.i.d. symbols, where in both cases we assume zero-mean proper Gaussian data symbols. When considering an optimized pilot-to-data power ratio, this gap becomes smaller but still exists. However, it has to be mentioned that this gap is relatively small in comparison to the achievable rate, indicating that for a SISO system the loss due to the restriction to synchronized detection in combination with a solely pilot based channel estimation is small.

Obviously, parts of the information leading to this gap might be exploitable while using receivers based on iterative synchronization and decoding. Therefore, we have studied the achievable rate with this type of receiver. These receivers iteratively enhance the channel estimation based on reliability information on the data symbols given by the decoder. We show that the possible gain of this arises due to the information contained in the temporal correlation of the channel estimation error in case of using a channel estimation which is solely based on pilot symbols. This information is discarded

while using synchronized detection, as the detector/decoder does not exploit the information contained in the correlation but treats the estimation error as a white process. Although receivers based on synchronized detection and an iterative code-aided channel estimation also use a symbol-wise detection metric, parts of the information contained in the temporal correlation of the channel estimation error when using a solely pilot based channel estimation can be retrieved by using an iterative code-aided channel estimation. Furthermore, we give an upper bound on the achievable rate for a specific iterative code-aided synchronized detection structure. In contrast to the typically studied receiver based on synchronized detection and iterative code-aided channel estimation, which can be motivated by expressing the joint ML detection and MAP parameter estimation problem based on a set of fixed point equations[2], we have given an upper bound on the achievable rate for a slightly modified receiver. The modification lies in the channel estimator, which does in general not use the observation y_k for the estimation of h_k. However, from a practical point of view, for low channel dynamics, we guess that the amount of hereby discarded information is rather small. Furthermore, we assume i.i.d. zero-mean proper Gaussian data symbols. We show that this receiver cannot exploit the complete mutual information between the transmitter and the receiver. The information corresponding to $\mathcal{I}(y_k; \mathbf{x}_{\setminus k}|\mathbf{y}_{\setminus k})$, which seems to be small for practical, i.e., small, channel dynamics, cannot be retrieved. The question if and which part of this information can be exploited without the modification to the channel estimator remains open. Unfortunately, the given upper bound on the achievable rate for the specific receiver based on iterative synchronized detection is a non-closed form expression, as it explicitly depends on the channel interpolation error variance at an arbitrarily chosen time instant k and, thus, is a function of the distribution of all past and all future transmit symbols with respect to this time instant. We have not found a closed form solution. However, for small channel dynamics it is reasonable to approximate the channel interpolation error variance by the channel interpolation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols. Based on this approximation we have numerically evaluated the upper bound on the achievable rate with the discussed iterative code-aided synchronized detection structure. Due to the approximation of the channel interpolation error variance, these results are only an approximation. The proof that this approach leads to an actual upper bound is missing. Opposed to this, for the case of constant modulus signaling, we can give a closed form solution for the upper bound on the achievable rate with the specific receiver based on synchronized detection and iterative code-aided channel estimation.

In contrast to the general upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs, for asymptotically small channel dynamics the upper bound on the achievable rate for the studied receiver using

---

[2] Notice, the same receiver can also be derived on the basis of the EM framework.

synchronized detection based on an iterative code-aided channel estimation and for i.i.d. zero-mean proper Gaussian data symbols converges to the coherent capacity. As the evaluation of this upper bound based on approximating the channel interpolation error variance with the interpolation error variance calculated under the assumption of constant modulus input symbols in the past and in the future is lower than the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for small channel dynamics, it gives a valuable upper bound on the achievable rate for the realistic iterative code-aided synchronized detection based receiver as it has been described before. The evaluation of this approximative upper bound for a fixed pilot spacing enables to give an approximate upper bound on the possible gain by using the specific iterative code-aided synchronized detection based receiver instead of a solely pilot based synchronized detection. As the amount of information that is discarded by the modification of the channel estimator seems to be small, the approximate upper bound also gives an indication for an upper bound on the possible gain by using iterative code-aided synchronized detection in general. A lower bound on the achievable rate with iterative synchronization and decoding is obviously given by the lower bound on the achievable rate with synchronization in combination with a solely pilot based channel estimation. However, as in this case the channel estimation is solely based on pilot symbols, it is not very tight.

As it has been discussed, receivers based on synchronized detection and an iterative code-aided channel estimation still use a symbol-wise detection metric, which is not optimal. Therefore, we also study the achievable rate with an optimal joint processing of pilot and data symbols, and give a lower bound on the achievable rate for the case of a stationary Rayleigh flat-fading channel. This lower bound can be seen as an extension of a lower bound on the achievable rate with joint processing of pilot and data symbols for the case of a block fading channel given in [56]. Furthermore, we compare this lower bound to the bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation and to the approximate upper bound on the achievable rate with the receiver based on synchronized detection and iterative code-aided channel estimation discussed in Chapter 6. We observe that the approximative upper bound on the achievable rate with the iterative synchronized detection based receiver is larger than the lower bound on the achievable rate with joint processing of pilot and data symbols. This is not self-evident as the receiver based on synchronized detection and iterative code-aided channel estimation discussed in Chapter 6 cannot exploit all information, i.e., the information corresponding to $\mathcal{I}(y_k; \mathbf{x}_{\setminus k}|\mathbf{y}_{\setminus k})$ cannot be retrieved by the receiver. In general the comparison of the lower bound on the achievable rate with joint processing of pilot and data symbols to the bounds on the achievable rate with synchronized detection using a solely pilot based channel estimation and to the upper bound for the receiver based on iterative code-aided channel estimation discussed in Chapter 6 gives an

indication on the possible gain when using enhanced receiver structures which also use data symbols for channel estimation.

In a further step, we extend the bounds on the achievable rate with i.d. inputs to the MIMO case, including spatial antenna correlation. As in the SISO case, the main innovation concerning these bounds is the fact that their derivation does not rely on a peak power constraint and, thus, enables bounding of the achievable rate with i.i.d. zero-mean proper Gaussian inputs. When evaluating these bounds for peak power constrained input symbols, they are closely related to bounds on capacity for the frequency selective MIMO channel given in [102]. Unfortunately, it has to be stated that these bounds become quite loose with an increasing number of receive antennas. This also obstructs general statements on the effect of spatial antenna correlation. Only for the special case of small channel dynamics, we see the same effects as they are observed for the case of perfect CSI at the receiver, i.e., antenna correlation at the receiver side decreases the achievable rate, while for antenna correlation at the transmitter side the effect depends on the SNR. For a high SNR antenna correlation decreases the achievable rate while it helps for low SNR.

In the context of the MIMO scenario, the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs to the achievable rate based on synchronized detection in combination with a solely pilot based channel estimation shows that for the important range of small channel dynamics the achievable rate with a pilot based synchronized detection is slightly lower than the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. However, due to the looseness of the bounds, it is not possible to give general statements on the performance loss in the achievable rate while restricting to pilot based synchronized detection.

Furthermore, we extend the bounds on the achievable rate with i.d. inputs to underspread frequency-selective channels and a brick-shaped scattering function, where we neglect intercarrier and intersymbol interference. As the derivation is not based on the assumption of a peak power constraint, these bounds can also be evaluated for proper Gaussian input symbols, which is the main novelty in this context. However, the upper bound holds only approximately, as an approximation on the eigenvalue distribution of the channel correlation matrix is required, which is not tight for finite bandwidths. In case of a peak power constraint our approximative upper bound on the achievable rate with i.d. input symbols is equivalent to the upper bound on capacity given in [33] and our lower bound can easily be modified by a time sharing argumentation yielding also the lower bound on capacity given in [33]. We compare the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to the achievable rate with synchronized detection in combination with a solely pilot based channel estimation and zero-mean proper Gaussian data symbols. Therefore, we use realistic channel parameters as they are encountered for IEEE 802.11a systems. This comparison shows that the achievable rate with a solely pilot based synchronized detection stays below the achievable rate with i.i.d. zero-mean proper Gaussian

input symbols with a gap that depends on the channel bandwidth and the channel dynamics. However, for typical system bandwidths the gap is relatively small in comparison to the achievable rate. If on the other hand the system is operated close to the critical bandwidth, the gap becomes significantly larger. For these comparisons it has to be kept in mind that the ISI and ICI has been neglected. Concerning future work, the study of the effect of interference onto the achievable rate is highly relevant.

Up to this point, we have considered the achievable rate with constraints on the input distribution like i.d. or i.i.d. input symbols. However, as it has been shown in [107] that at least i.i.d. input symbols are not optimal in general, we deviate from this assumption. Therefore, for a fixed set of discrete input sequences and a time-selective block Rayleigh flat-fading channel, we have studied the input probability distribution, i.e., the probability of occurrence of the individual input sequences, that achieves the capacity constrained to the given input set. The discrete input distribution that achieves the constrained capacity is characterized by a constant Kullback-Leibler distance between the output PDFs conditioned on the individual input sequences and the unconditional output PDF. For the special case of PSK signaling the optimum input distribution corresponds to a uniform input distribution over all distinguishable input sequences. Furthermore, the special case of using one specific time instant for a pilot symbol and i.i.d. data symbols at the other time instances is included in the set of constrained capacity-achieving input distributions. In addition, for the time-selective block fading channel with block length N the asymptotic high SNR constrained capacity is degraded at least by a factor of $\frac{N-1}{N}$ with respect to the case of perfect channel state information at the receiver. Concerning this work, it would be interesting to find the constrained capacity-achieving input distributions for more general modulation schemes including amplitude modulation, i.e., QAM. In addition, it is known that for the considered channel the capacity-achieving input distribution itself is discrete with a finite number of mass points [10]. Therefore, optimization with respect to the optimum number of signaling points as well as with respect to their localization and their distribution is an important open problem.

The finding that for time-selective block fading and PSK modulation the use of a single pilot symbol per block and i.i.d. data symbols at all other time instances is optimal also means that, at least for PSK modulation, periodically inserted pilot symbols are not optimal with respect to the achievable rate in the context of stationary fading. Nevertheless, the use of periodic pilot symbols is a smart approach, as they allow for a high spectral efficiency in combination with receivers of low computational complexity. Besides of this, also different alternatives have been proposed in literature to get to a noncoherent decoding with suitable complexity.

In conclusion, this work has delivered some new bounds and concepts concerning the understanding of the capacity/achievable rate of stationary fading channels and their relation to the achievable rate with synchronized

detection and iterative synchronization and decoding. However, the general question on the capacity and the capacity-achieving input distribution of stationary fading channels, which turns out to be notoriously difficult to answer, remains open.

Appendix A
Mathematical Derivations and Proofs

A.1 Modified Upper Bound on $h'(\mathbf{y})$ for PG Inputs

In this appendix, we derive an alternative upper bound on the channel output entropy rate $h'(\mathbf{y})$ for the special case of zero-mean proper Gaussian input symbols, which is tighter than the one given in (3.18). This derivation is based on work given in [88], [87]. On the other hand its evaluation requires some more complex numerical methods. Therefore, we do not further use this bound. For completeness of presentation, we give it within this appendix.

Obviously, an upper bound on the entropy $h(\mathbf{y})$ is given by assuming an uncorrelated channel fading process, i.e., its correlation matrix is assumed to be diagonal. While assuming an uncorrelated channel fading process, more randomness is introduced into the channel output vector $\mathbf{y}$, leading to an increased entropy. Furthermore, an uncorrelated channel leads to independent channel observations. Thus, $h'(\mathbf{y})$ is upper-bounded by

$$h'(\mathbf{y}) \leq h(y_k). \tag{A.1}$$

The major difference between this upper bound and the upper bound given in (3.18) is that the latter one implicitly corresponds to the case that the channel observations y_k are proper Gaussian, while the RHS of (A.1) still corresponds to the actual channel output entropy of the individual time instances. The upper bounding in (A.1) only discards the temporal dependencies between the different observations.

In the following we calculate the entropy $h(y_k)$ for the specific case of zero-mean proper Gaussian input symbols with an average power σ_x^2

$$h(y_k) = -\mathrm{E}_{y_k}\left[\log(p(y_k))\right]$$
$$= -\int_{\mathbb{C}}\int_{\mathbb{C}} p(y_k|x_k)p(x_k)dx_k \log\left(\int_{\mathbb{C}} p(y_k|x_k)p(x_k)dx_k\right)dy_k$$

$$= -\int_0^\infty \left[\int_0^\infty \frac{2|y|}{\sigma_h^2 |x|^2 + \sigma_n^2} e^{-\frac{|y|^2}{\sigma_h^2 |x|^2 + \sigma_n^2}} \frac{2|x|}{\sigma_x^2} e^{-\frac{|x|^2}{\sigma_x^2}} d|x| \right]$$

$$\times \log \left(\int_0^\infty \frac{2|y|}{\sigma_h^2 |x|^2 + \sigma_n^2} e^{-\frac{|y|^2}{\sigma_h^2 |x|^2 + \sigma_n^2}} \frac{2|x|}{\sigma_x^2} e^{-\frac{|x|^2}{\sigma_x^2}} d|x| \right) d|y|$$

$$+ \log(2\pi) - \frac{\gamma}{2} + \frac{1}{2} \int_{z=0}^\infty \log \left(\sigma_h^2 \sigma_x^2 z + \sigma_n^2 \right) e^{-z} dz \tag{A.2}$$

$$= h'_{U_2}(\mathbf{y}) \tag{A.3}$$

where $\gamma \approx 0.57721$ is the Euler constant. To the best of our knowledge the first integral in (A.2) cannot be calculated analytically. However, it can be evaluated numerically using Hermite polynomials and Simpson's rule, see [88], [87], [116], or by Monte Carlo integration.

For the evaluation of the tightness of $h'_{U_2}(\mathbf{y})$, in Fig. A.1 the difference

$$\Delta_{h'(\mathbf{y}),2} = h'_{U_2}(\mathbf{y}) - h'_L(\mathbf{y})\big|_{\mathrm{PG}} \tag{A.4}$$

is shown in comparison to the difference $\Delta_{h'(\mathbf{y})}$ given in (3.99). Obviously, the upper bound $h'_{U_2}(\mathbf{y})$ is tighter than the upper bound $h'_U(\mathbf{y})$ given in (3.18).
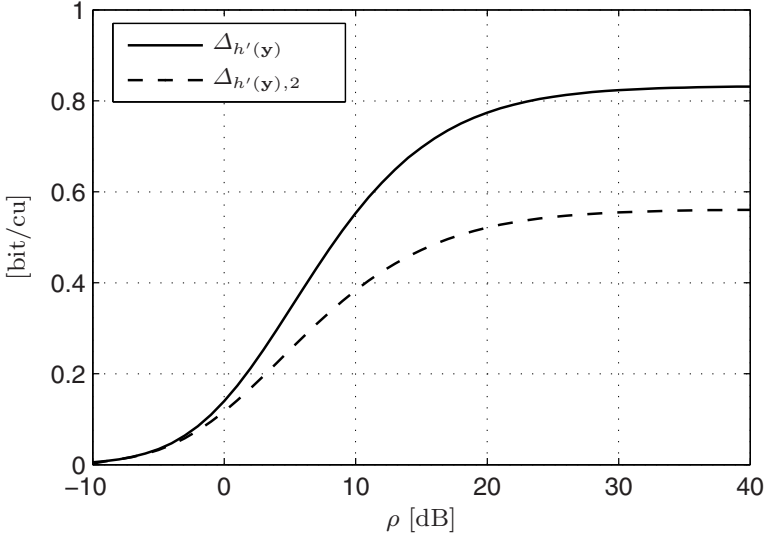
Fig. A.1: Comparison of $\Delta_{h'(\mathbf{y}),2}$ with $\Delta_{h'(\mathbf{y})}$

A.2 Calculation of Sufficient Conditions for $\alpha_{\mathrm{opt}} = 1$ in (3.76)

In this appendix, we give conditions on the parameters f_d, ρ, and β so that $\alpha_{\mathrm{opt}} = 1$ in (3.76), i.e., the upper bound in (3.74) is maximized by choosing the maximum average power σ_x^2. Therefor, we have to evaluate for which parameter choice the following inequality holds

$$\left(\frac{2f_d}{\beta} \log \left(\frac{\rho\beta}{2f_d} + 1 \right) \right)^{-1} - \frac{1}{\rho} \geq 1$$

$$\Leftrightarrow \frac{2f_d}{\beta} \log \left(\frac{\rho\beta}{2f_d} + 1 \right) \leq \frac{\rho}{1 + \rho}. \tag{A.5}$$

The following calculations are closely related to a corresponding problem in [33, Appendix C]. We divide the evaluation into the two cases $\rho > 1$ and $\rho \leq 1$.

For $\rho > 1$ the RHS of (A.5) can be lower-bounded by

$$\frac{\rho}{1 + \rho} \geq \frac{1}{2} \tag{A.6}$$

yielding the following sufficient condition for (A.5) to hold

$$\frac{2f_d}{\beta} \log \left(\frac{\rho\beta}{2f_d} + 1 \right) \leq \frac{1}{2}$$

$$\Leftrightarrow \rho \leq \frac{2f_d}{\beta} \left[\exp \left(\frac{1}{2} \frac{\beta}{2f_d} \right) - 1 \right]. \tag{A.7}$$

Thus, $\alpha_{\mathrm{opt}} = 1$ holds if

$$1 < \rho \leq \frac{2f_d}{\beta} \left[\exp \left(\frac{1}{2} \frac{\beta}{2f_d} \right) - 1 \right]. \tag{A.8}$$

Now, we discuss the case $\rho \leq 1$. Using the inequality $\frac{1}{x} \log(x + 1) \leq \frac{1}{\sqrt{x+1}}$ for $x \geq 0$, for $\rho \leq 1$ the LHS of (A.5) can be upper-bounded by

$$\frac{2f_d}{\beta} \log \left(\frac{\rho\beta}{2f_d} + 1 \right) \leq \frac{\rho}{\sqrt{\frac{\rho\beta}{2f_d} + 1}}. \tag{A.9}$$

Based on (A.9), inequality (A.5) holds if the following sufficient condition is fulfilled

$$\frac{\rho}{\sqrt{\frac{\rho\beta}{2f_d}+1}} \leq \frac{\rho}{1+\rho}$$

$$\Leftrightarrow 2f_d \leq \frac{\beta}{\rho+2} \tag{A.10}$$

so that we get the second condition

$$2f_d \leq \frac{\beta}{\rho+2} \text{ for } \rho \leq 1. \tag{A.11}$$

Thus, if (A.8) or (A.11) is fulfilled, (3.76) yields $\alpha_{\mathrm{opt}}=1$.

A.3 Proof of Monotonicity of $\Delta_{h'(\mathbf{y})}$

In this section, we prove the monotonicity of the difference $\Delta_{h'(\mathbf{y})}$ given in (3.99).

Proof. The difference can be transformed as follows

$$\Delta_{h'(\mathbf{y})} = \log\left(\rho+1\right) - \int_0^\infty \log\left(\rho z + 1\right) e^{-z} dz$$

$$= \int_0^\infty \log\left(\frac{\rho+1}{\rho z+1}\right) e^{-z} dz. \tag{A.12}$$

To show monotonicity, we take the derivative of $\Delta_{h'(\mathbf{y})}$ with respect to the SNR ρ yielding

$$\frac{\partial \Delta_{h'(\mathbf{y})}}{\partial \rho} = \int_0^\infty \frac{\partial \log\left(\frac{\rho+1}{\rho z+1}\right)}{\partial \rho} e^{-z} dz$$

$$= \frac{1}{\rho+1} \int_0^\infty \frac{1-z}{\rho z+1} e^{-z} dz. \tag{A.13}$$

As the factor in front of the integral is positive, $\Delta_{h'(\mathbf{y})}$ monotonically increases if

$$\int_0^\infty \frac{1-z}{\rho z+1} e^{-z} dz \geq 0. \tag{A.14}$$

Next, we show that this condition is fulfilled. Therefore, we rewrite the LHS of (A.14) as

$$\int_0^\infty \frac{1-z}{\rho z+1}e^{-z}dz = \int_0^\infty \frac{1+\frac{1}{\rho}-\frac{1}{\rho}(\rho z+1)}{\rho z+1}e^{-z}dz$$

$$= \int_0^\infty \frac{1+\frac{1}{\rho}}{\rho z+1}e^{-z}dz - \frac{1}{\rho}\int_0^\infty e^{-z}dz$$

$$= \frac{1}{\rho}\int_0^\infty \frac{1+\rho}{\rho z+1}e^{-z}dz - \frac{1}{\rho}$$

$$= \frac{1}{\rho}\left[\int_0^\infty \frac{1+\rho}{\rho z+1}e^{-z}dz - 1\right]. \tag{A.15}$$

The condition in (A.14) is fulfilled if

$$\int_0^\infty \frac{1+\rho}{\rho z+1}e^{-z}dz \geq 1. \tag{A.16}$$

In the following we show that (A.16) holds. Therefore, we evaluate the behavior of the integrand in (A.16) given by $f(z)e^{-z}$ with

$$f(z) = \frac{1+\rho}{1+\rho z}. \tag{A.17}$$

The goal is to lower-bound $f(z)$, enabling to show that (A.16) holds. The function $f(z)$ monotonically decreases and is convex in z, as

$$\frac{\partial f(z)}{\partial z} = -\frac{\rho(1+\rho)}{(1+\rho z)^2} \leq 0 \tag{A.18}$$

$$\frac{\partial^2 f(z)}{\partial z^2} = \frac{2\rho^2(1+\rho)}{(1+\rho z)^3} \geq 0 \qquad \forall z \in [0,\infty[. \tag{A.19}$$

Evaluation of $f(z)$ for the following values

$$f(0) = 1+\rho \tag{A.20}$$

$$f(1) = 1 \tag{A.21}$$

$$\lim_{z\to\infty} f(z) = 0 \tag{A.22}$$

shows that $f(z) > 0$ for $z \geq 0$. We now lower bound $f(z)$ by a tangent $f_1(z)$ to $f(z)$ in the point $z = 1$, which is given by

$$f_1(z) = \frac{1+2\rho}{1+\rho} - \frac{\rho}{1+\rho}z. \tag{A.23}$$

As in addition $f(z)$ is positive, we can lower-bound it by 0. Thus, the integral in (A.16) is lower-bounded by

$$\int_0^\infty \frac{1+\rho}{\rho z+1}e^{-z}dz \geq \int_0^{z_0} f_1(z)e^{-z}dz \tag{A.24}$$

where z_0 is the argument for which the tangent $f_1(z)$ becomes zero, i.e., $f_1(z_0) = 0$, and it is given by

$$z_0 = \frac{1 + 2\rho}{\rho}. \tag{A.25}$$

Thus, the RHS of (A.24) is given by

$$\int_0^{z_0} f_1(z)e^{-z}dz = \int_0^{\frac{1+2\rho}{\rho}} \left(\frac{1 + 2\rho}{1 + \rho} - \frac{\rho}{1 + \rho}z\right) e^{-z}dz$$

$$= 1 + \frac{\rho}{1 + \rho} \exp\left(-\frac{1 + 2\rho}{\rho}\right)$$

$$\geq 1. \tag{A.26}$$

Hence, we have shown that the condition given in (A.16) is fulfilled and, thus, $\Delta_{h'(\mathbf{y})}$ is monotonically increasing. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

A.4 Calculation of $\mathrm{E}[\varepsilon_{\mathrm{pred}}]$ for the Enhanced Upper Bound on $h'(\mathbf{y})$

In this appendix, we calculate the mean of the channel power prediction error $\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right]$ of the estimator defined in Section 4.2.2 in the context of the enhanced upper bound on the channel output entropy rate $h'(\mathbf{y})$. Furthermore, we give closed form expressions for $\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right]$ for the special case that the channel fading process is characterized by a rectangular PSD.

We want to calculate $\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right]$ which corresponds to $\mathrm{E}\left[\varepsilon_N\right]$ for an infinite number of observations in the past, i.e.,

$$\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right] = \lim_{N \to \infty} \mathrm{E}\left[\varepsilon_N\right]. \tag{A.27}$$

Recalling (4.35), $\mathrm{E}\left[\varepsilon_N\right]$ is given by

$$\mathrm{E}\left[\varepsilon_N\right] = \sigma_h^2 \left(1 - \sum_{l=1}^{N-1} w_l\right). \tag{A.28}$$

For the calculation of $\mathrm{E}\left[\varepsilon_N\right]$, we need to discuss the behavior of the filter coefficients w_l contained in $\mathbf{w}$, i.e., of the LMMSE predictor defined by (4.30) and (4.29). Note that the filter coefficients contained in $\mathbf{w}$ are the filter coefficients, yielding an LMMSE estimate of $\widetilde{|y_N|^2}$ based on the zero-mean samples $\widetilde{|y_1|^2}, \ldots, \widetilde{|y_{N-1}|^2}$. Obviously, (A.28) is the difference between the constant σ_h^2 and the output of the filter given by the coefficients in $\mathbf{w}$ when the constant σ_h^2 is at its input. Obviously, it corresponds to the estimation error of the

predictor in case a constant lies at its input. Note that here the filter coefficients $\mathbf{w}$ are chosen corresponding to the dynamics of the channel, and not for a constant channel.

For the discussion of (A.28) first consider the case that we want to predict the value a_N at the time instance N of a zero-mean process $\{a_k\}$ based on an infinite number of observations in the past, i.e., $N \to \infty$. If $\hat{a}_N$ is the prediction of a_N, the prediction error is given by

$$e_{a_N} = a_N - \hat{a}_N. \tag{A.29}$$

Let $H_{\mathrm{pred}}(f)$ be the innovation filter yielding e_{a_N} based on $a_1, \ldots, a_{N-1}$ and $N \to \infty$.

For $N \to \infty$, the prediction error variance of the one-step predictor is given by, see [44, Chapter 10.8] and cf. (A.67)

$$\sigma^2_{e_{a,\mathrm{pred}}} = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(S_a(f)\right) df\right) \tag{A.30}$$

where $S_a(f)$ is the PSD of the filter input process. Due to the fact that the prediction error process $\{e_{a_k}\}$ is white the following relation holds [90]

$$\sigma^2_{e_{a,\mathrm{pred}}} = S_a(f) \left|H_{\mathrm{pred}}(f)\right|^2. \tag{A.31}$$

Thus, based on (A.30) and (A.31) we obtain the following implicit expression for the prediction innovation filter $H_{\mathrm{pred}}(f)$

$$\left|H_{\mathrm{pred}}(f)\right|^2 = \frac{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(S_a(f)\right) df\right)}{S_a(f)}. \tag{A.32}$$

Identifying $\widetilde{|y_N|^2}$ with a_N and $\widehat{\widetilde{|y_N|^2}}$ with $\hat{a}_N$, the difference on the RHS of (A.28) corresponds to the output of the innovation filter $H_{\mathrm{pred}}(f)$ in case the constant σ_h^2 lies at its input. As the input signal is a constant, we have to evaluate the innovation filter for $f = 0$, and (A.28) becomes

$$\mathrm{E}\left[\varepsilon_{\mathrm{pred}}\right] \equiv \lim_{N \to \infty} \mathrm{E}\left[\varepsilon_N\right] = \sigma_h^2 H_{\mathrm{pred}}(0). \tag{A.33}$$

Based on (A.32), $H_{\mathrm{pred}}(0)$ is given by

$$H_{\mathrm{pred}}(0) = \sqrt{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_{\widetilde{|y|^2}}(f)}{S_{\widetilde{|y|^2}}(0)}\right) df\right)} \tag{A.34}$$

where $S_{\widetilde{|y|^2}}(f)$ is the PSD of the process $\left\{\widetilde{|y|^2}\right\}$.

For the determination of $S_{\widetilde{|y|^2}}(f)$, we calculate its autocorrelation function first. The autocorrelation function of the process $\left\{\widetilde{|y_k|^2}\right\}$ is given by

$$
\begin{aligned}
r_{\widetilde{|y|^2}}(l) &= \mathrm{E}\left[\widetilde{|y_k|^2}\,\widetilde{|y_{k+l}|^2}\right] \\
&= \mathrm{E}\left[|y_k|^2|y_{k+l}|^2\right] - \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)^2 \\
&= \mathrm{E}\left[|x_k h_k + n_k|^2|x_{k+l}h_{k+l} + n_{k+l}|^2\right] - \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)^2 \\
&= \begin{cases} 4\alpha^2\sigma_x^4\sigma_h^4 + 4\alpha\sigma_x^2\sigma_h^2\sigma_n^2 + 2\sigma_n^4 - \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)^2 & \text{for } l = 0 \\ \alpha^2\sigma_x^4\mathrm{E}\left[|h_k|^2|h_{k+l}|^2\right] + 2\alpha\sigma_x^2\sigma_h^2\sigma_n^2 + \sigma_n^4 - \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)^2 & \text{for } l \neq 0 \end{cases} \\
&\overset{(a)}{=} \begin{cases} 3\alpha^2\sigma_x^4\sigma_h^4 + 2\alpha\sigma_x^2\sigma_h^2\sigma_n^2 + \sigma_n^4 & \text{for } l = 0 \\ \alpha^2\sigma_x^4[r_h(l)]^2 & \text{for } l \neq 0 \end{cases} \\
&= \alpha^2\sigma_x^4[r_h(l)]^2 + \left(2\alpha^2\sigma_x^4\sigma_h^4 + 2\alpha\sigma_x^2\sigma_h^2\sigma_n^2 + \sigma_n^4\right)\delta_l \qquad\qquad\text{(A.35)}
\end{aligned}
$$

where δ_l is the Kronecker delta and where $r_h(l)$ is the autocorrelation function of the channel fading process $\{h_k\}$ as defined in (2.2). For (a) we used the fact that

$$
\mathrm{E}\left[|h_k|^2|h_{k+l}|^2\right] = \sigma_h^4 + [r_h(l)]^2. \qquad\qquad\text{(A.36)}
$$

Based on the autocorrelation function $r_{\widetilde{|y|^2}}(l)$ in (A.35) we can calculate the PSD of the process $\left\{\widetilde{|y_k|^2}\right\}$ which is given by

$$
\begin{aligned}
S_{\widetilde{|y|^2}}(f) &= \sum_{l=-\infty}^{\infty} r_{\widetilde{|y|^2}}(l)e^{-j2\pi l f} \\
&= \alpha^2\sigma_x^4 \sum_{l=-\infty}^{\infty} \left\{[r_h(l)]^2 e^{-j2\pi l f}\right\} + 2\alpha^2\sigma_x^4\sigma_h^4 + 2\alpha\sigma_x^2\sigma_h^2\sigma_n^2 + \sigma_n^4.
\end{aligned}
$$

$$\text{(A.37)}$$

For the special case that the channel fading process is characterized by a rectangular PSD, the corresponding time-continuous autocorrelation function is given by, cf. (2.9)

$$
r_h(t) = \sigma_h^2 \mathrm{sinc}\left(2f_d \frac{t}{T_{\mathrm{Sym}}}\right). \qquad\qquad\text{(A.38)}
$$

The spectrum of the time-continuous autocorrelation function corresponding to $[r_h(l)]^2$ as it occurs in (A.37) is given by

$$
\bar{S}_{|h|^2}(f) = \frac{\sigma_h^4 T_{\mathrm{Sym}}}{2f_d}\mathrm{tri}\left(\frac{f}{2f_d}\right) \qquad\qquad\text{(A.39)}
$$

where tri indicates the triangular function given by

$$\text{tri}(t) = \begin{cases} 1 - |t| & \text{for } |t| < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{A.40}$$

Thus, the PSD corresponding to the discrete-time process $\{|h_k|^2\}$ is given by

$$S_{|h|^2}(f) = \sum_{k=-\infty}^{\infty} \frac{\sigma_h^4}{2f_d} \text{tri}\left(\frac{f-k}{2f_d}\right). \tag{A.41}$$

Now we distinguish the following two cases

- $f_d < \frac{1}{4}$, i.e., the case that $S_{|h|^2}(f)$ consists of non-overlapping replicas,
- $\frac{1}{4} \leq f_d \leq \frac{1}{2}$, i.e., the replicas are overlapping.

For these two cases we get the following PSDs in the interval $f \in [-\frac{1}{2}, \frac{1}{2}]$

$$S_{|\widetilde{y}|^2}(f)\Big|_{f_d < \frac{1}{4}} = \frac{\alpha^2 \sigma_x^4 \sigma_h^4}{4f_d^2} \max\{2f_d - |f|, 0\} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2 + \sigma_n^4 \tag{A.42}$$

$$S_{|\widetilde{y}|^2}(f)\Big|_{\frac{1}{4} \leq f_d \leq \frac{1}{2}} = \frac{\alpha^2 \sigma_x^4 \sigma_h^4}{4f_d^2} \max\{2f_d - |f|, 4f_d - 1\} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2$$
$$+ \sigma_n^4. \tag{A.43}$$

Thus, for the special case of a rectangular PSD of the channel fading process we get with (A.33) and (A.34) the following E[$\varepsilon_{\text{pred}}$]

$$E[\varepsilon_{\text{pred}}]\Big|_{f_d < \frac{1}{4}}$$
$$= \sigma_h^2 \exp\left\{ \frac{1}{2} \left(1 - 2\frac{4f_d^2}{\alpha^2 \sigma_x^4 \sigma_h^4} \left(\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{2f_d} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2 + \sigma_n^4\right)\right)\right.$$
$$\left. \times \log\left(1 - \frac{\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{2f_d}}{\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{2f_d} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2 + \sigma_n^4}\right) - 2f_d\right\} \tag{A.44}$$

$$E[\varepsilon_{\text{pred}}]\Big|_{\frac{1}{4} \leq f_d \leq \frac{1}{2}}$$
$$= \sigma_h^2 \exp\left\{ \frac{1}{2} \left(1 - 2\frac{4f_d^2}{\alpha^2 \sigma_x^4 \sigma_h^4} \left(\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{2f_d} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2 + \sigma_n^4\right)\right)\right.$$
$$\left. \times \log\left(1 - \frac{(1 - 2f_d)\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{4f_d^2}}{\frac{\alpha^2 \sigma_x^4 \sigma_h^4}{2f_d} + 2\alpha^2 \sigma_x^4 \sigma_h^4 + 2\alpha \sigma_x^2 \sigma_h^2 \sigma_n^2 + \sigma_n^4}\right) - (1 - 2f_d)\right\}. \tag{A.45}$$

These expressions for $\mathrm{E}[\varepsilon_{\mathrm{pred}}]$ hold for the special case of a rectangular PSD. It can easily be checked, that (A.44) and (A.45) fulfill the conditions given in (4.26) and (4.27).

A.5 Proof of Convexity of (4.49)

In this appendix, we will show that the expression at the RHS of (4.49), i.e.,

$$
\log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\sigma_{e_{\mathrm{pred}}}^2\left(\mathbf{x}_1^{N-1}\right)\right)
$$

$$
= \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\mathrm{pred}}\right)\right) \quad \text{(A.46)}
$$

is convex with respect to the individual elements of the diagonal matrix $\mathbf{Z}$.

For this purpose, in a first step we rewrite the channel prediction error variance as follows

$$
\sigma_{e_{\mathrm{pred}}}^2\left(\mathbf{x}_1^{N-1}\right) = \sigma_{e_{\mathrm{pred}}}^2(\mathbf{z}) = \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h + \sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\mathrm{pred}}
$$

$$
\overset{(a)}{=} \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left(\frac{1}{\sigma_n^2}\mathbf{Z} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\mathrm{pred}}
$$

$$
\overset{(b)}{=} \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left(\frac{1}{\sigma_n^2}\left(z_i\mathbf{V}_i + \mathbf{Z}_{\backslash i}\right) + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\mathrm{pred}}
$$

$$
= \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left[\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1}\right)\right.\right.
$$

$$
\left.\left.\times\left\{\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1}\right)^{-1}\frac{z_i}{\sigma_n^2}\mathbf{V}_i + \mathbf{I}\right\}\right]^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\mathrm{pred}}
$$

$$
= \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left\{\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1}\right)^{-1}\frac{z_i}{\sigma_n^2}\mathbf{V}_i + \mathbf{I}\right\}^{-1}\right.
$$

$$
\left.\times\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\mathrm{pred}}
$$

$$
\overset{(c)}{=} \sigma_h^2 - \mathbf{r}_{h,\mathrm{pred}}^H\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left\{\mathbf{I} - \frac{z_i}{1 + z_i\lambda_{\max}}\left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1}\right)^{-1}\frac{\mathbf{V}_i}{\sigma_n^2}\right\}\right.
$$

$$
\left.\times\left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\mathrm{pred}}
$$

$$= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}}$$

$$- \frac{z_i}{1 + z_i \lambda_{\max}} \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}$$

$$\overset{(d)}{=} \sigma_{e_{\text{pred}}}^2 (\mathbf{z}_{\backslash i})$$

$$- \frac{z_i}{1 + z_i \lambda_{\max}} \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left(\frac{\mathbf{Z}_{\backslash i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}$$

$$\overset{(e)}{=} \sigma_{e_{\text{pred}}}^2 (\mathbf{z}_{\backslash i}) - \frac{z_i a}{1 + z_i \lambda_{\max}} \tag{A.47}$$

where for (a) we have used the matrix inversion lemma. For (b) we have separated the diagonal matrix $\mathbf{Z}$ as follows

$$\mathbf{Z} = \mathbf{Z}_{\backslash i} + z_i \mathbf{V}_i \tag{A.48}$$

where $\mathbf{Z}_{\backslash i}$ corresponds to $\mathbf{Z}$ except that the i-th diagonal element is set to 0, $\mathbf{V}_i$ is a matrix with all elements zero except of the i-th diagonal element being equal to the 1, and z_i is the i-th diagonal element of the matrix $\mathbf{Z}$. For (c) we have used that the matrix

$$z_i \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i = z_i \mathbf{B} \tag{A.49}$$

is of rank 1 and, therefore, the following equality holds (*Sherman-Morrison formula*)

$$(\mathbf{I} + z_i \mathbf{B})^{-1} = \mathbf{I} - \frac{z_i}{1 + z_i \lambda_{\max}} \mathbf{B} \tag{A.50}$$

where $\lambda_{\max}$ is the non-zero eigenvalue of $\mathbf{B}$.

Furthermore, (d) is based on the definition

$$\sigma_{e_{\text{pred}}}^2 (\mathbf{z}_{\backslash i}) = \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \tag{A.51}$$

which is the prediction error variance if the observation at the i-th time instant is not used for the channel prediction. This becomes obvious by comparison with (A.47(a)). Finally for (e) we have used the substitution

$$a = \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}$$

$$\geq 0 \tag{A.52}$$

where the nonnegativity follows from the fact that $\mathbf{V}_i$ is positive semidefinite.

Thus, with (A.47) we have found a separation of the channel prediction error variance $\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z})$ into the term $\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z}_{\setminus i})$ being independent of z_i, and an additional term, which depends on z_i. Note that a and $\lambda_{\max}$ in the second term on the RHS of (A.47) are independent of z_i. Note that the element i is an arbitrary chosen element, i.e., we can use this separation for each diagonal element of the matrix $\mathbf{Z}$.

By substituting the RHS of (A.47) into (A.46) we get

$$
\log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\sigma^2_{e_{\mathrm{pred}}}(\mathbf{x}_1^{N-1})\right)
$$
$$
= \log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z}_{\setminus i}) - \frac{z_i a}{1 + z_i\lambda_{\max}}\right)\right) = K. \tag{A.53}
$$

Recall that we want to show the convexity of (A.53) with respect to the element z_i. Therefore, we calculate its second derivative with respect to z_i:

$$
\frac{\partial K}{\partial z_i} = -\frac{\frac{|x_N|^2}{\sigma_n^2}\frac{a}{(1+z_i\lambda_{\max})^2}}{1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z}_{\setminus i}) - \frac{z_i a}{1+z_i\lambda_{\max}}\right)}
$$

$$
\frac{\partial^2 K}{(\partial z_i)^2} = \frac{\frac{|x_N|^2}{\sigma_n^2}\frac{a\cdot 2\lambda_{\max}(1+z_i\lambda_{\max})}{(1+z_i\lambda_{\max})^4}}{\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a z_i}{1+z_i\lambda_{\max}}\right)\right)^2}
$$
$$
\times \left\{1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\mathrm{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a}{1+z_i\lambda_{\max}}\left(z_i + \frac{1}{2\lambda_{\max}}\right)\right)\right\}. \tag{A.54}
$$

Now we show that the second derivative is nonnegative, i.e.,

$$
\frac{\partial^2 K}{(\partial z_i)^2} \geq 0. \tag{A.55}
$$

Therefore, first we will show that $\lambda_{\max}$ is nonnegative. This can be easily done based on the definition of the eigenvalues of the matrix $\mathbf{B}$

$$
\mathbf{B}\mathbf{u} = \lambda_{\max}\mathbf{u}
$$
$$
\Leftrightarrow \left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)^{-1}\frac{1}{\sigma_n^2}\mathbf{V}_i\mathbf{u} = \lambda_{\max}\mathbf{u}
$$
$$
\Leftrightarrow \frac{1}{\sigma_n^2}\mathbf{V}_i\mathbf{u} = \left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)\lambda_{\max}\mathbf{u}
$$
$$
\Leftrightarrow \frac{1}{\sigma_n^2}\mathbf{u}^H\mathbf{V}_i\mathbf{u} = \lambda_{\max}\mathbf{u}^H\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)\mathbf{u}
$$
$$
\stackrel{(a)}{\Rightarrow} \lambda_{\max} \geq 0 \tag{A.56}
$$

where (a) follows from the fact that the eigenvalues of $\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\backslash i} + \mathbf{R}_h^{-1}\right)$ are nonnegative, as $\mathbf{R}_h$ is positive definite and the diagonal entries of the diagonal matrix $\mathbf{Z}_{\backslash i}$ are also nonnegative. In addition, obviously $\mathbf{V}_i$ is also positive semidefinite.

With $\lambda_{\max}$, z_i, and a being nonnegative, for the proof of (A.55) it rest to show that

$$\sigma_{e_{\mathrm{pred}}}^2(\mathbf{z}_{\backslash i}) - \frac{a}{1 + z_i\lambda_{\max}}\left(z_i + \frac{1}{2\lambda_{\max}}\right) \geq 0. \tag{A.57}$$

To prove this inequality we calculate the derivative of the LHS of (A.57) with respect to z_i, which is given by

$$\frac{\partial}{\partial z_i}\left\{\sigma_{e_{\mathrm{pred}}}^2(\mathbf{z}_{\backslash i}) - \frac{a}{1 + z_i\lambda_{\max}}\left(z_i + \frac{1}{2\lambda_{\max}}\right)\right\} = -\frac{a}{2(1 + z_i\lambda_{\max})^2}$$
$$\leq 0 \tag{A.58}$$

where for the last inequality we have used (A.52). I.e., the LHS of (A.57) monotonically decreases in z_i. Furthermore, for $z_i \to \infty$ the LHS of (A.57) becomes

$$\lim_{z_i\to\infty}\left\{\sigma_{e_{\mathrm{pred}}}^2(\mathbf{z}_{\backslash i}) - \frac{a}{1 + z_i\lambda_{\max}}\left(z_i + \frac{1}{2\lambda_{\max}}\right)\right\}$$
$$= \lim_{z_i\to\infty}\left\{\sigma_{e_{\mathrm{pred}}}^2(\mathbf{z}_{\backslash i}) - \frac{a}{1 + z_i\lambda_{\max}}z_i\right\}$$
$$\overset{(a)}{=} \lim_{z_i\to\infty}\sigma_{e_{\mathrm{pred}}}^2(\mathbf{z})$$
$$\overset{(b)}{\geq} 0 \tag{A.59}$$

where (a) follows due to (A.47), and where (b) holds as the prediction error variance must be nonnegative. As the LHS of (A.57) is monotonically decreasing in z_i and as its limit for $z_i \to \infty$ is nonnegative, (A.57) must hold.

Thus, with (A.57) inequality (A.55) holds, i.e., the second derivative of (A.53) with respect to z_i is nonnegative and, thus, (A.53) and (A.46) is convex in z_i.

In conclusion, we have shown that (A.46) is convex in each z_i for $i = 1, \ldots, N - 1$.

A.6 One-Step Prediction Error Variance

In this appendix, we calculate the prediction error variance

$$\sigma_e^2 = \mathrm{E}\left[|h_N - \hat{h}_N|^2\right] \tag{A.60}$$

where $\hat{h}_N$ is the LMMSE estimate of h_N based on an infinite number of noisy observations in the past

$$y_k = h_k + n_k, \quad k = 1, \ldots, N-1. \tag{A.61}$$

The prediction error variance σ_e^2 for this case is known, see e.g., [67]. For completeness of presentation, here we give a derivation that is closely related to the calculation of the estimation error variance in the context of *filtering* stationary sequences in white noise in [129]. The difference between the problem discussed in the present work and the one discussed in [129] is that in the latter one also an observation y_N is available and used for estimation of h_N. Therefore, this problem is titled *filtering* in contrast to the *prediction* considered here.

Following the approach in [129], the prediction error variance is given by

$$\sigma_e^2 = \frac{\det(\mathbf{B})}{\det\left(\mathbf{R}_{y,N-1}\right)} \tag{A.62}$$

where

$$\mathbf{R}_{y,N-1} = \mathbf{R}_{h,N-1} + \sigma_n^2 \mathbf{I}_{N-1} \quad \in \mathcal{C}^{N-1 \times N-1} \tag{A.63}$$

is the autocorrelation matrix of the past $N-1$ observation. Here the index $N-1$ indicates the size of the matrix $\mathbf{R}_{y,N-1}$. $\mathbf{R}_{h,N-1}$ corresponds to (2.17) but with dimension $N-1 \times N-1$.

The matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{pmatrix} \mathbf{R}_{y,N-1} & \mathbf{b}^H \\ \mathbf{b} & r_h(0) \end{pmatrix} \tag{A.64}$$

where

$$\mathbf{b} = [r_h(N-1), r_h(N-2), \ldots, r_h(2), r_h(1)]. \tag{A.65}$$

The estimation error variance (A.62) can be transformed to

$$\sigma_e^2 = \frac{\det\left(\mathbf{R}_{y,N}\right)}{\det\left(\mathbf{R}_{y,N-1}\right)} - \sigma_n^2. \tag{A.66}$$

For $N \to \infty$ the ratio of the determinants in (A.66) becomes, [118, Satz XVII], [44, Chapter 5.2, eq. (13)]

$$\lim_{N \to \infty} \frac{\det\left(\mathbf{R}_{y,N}\right)}{\det\left(\mathbf{R}_{y,N-1}\right)} = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\sigma_n^2 + S_h(f)\right) df\right) \tag{A.67}$$

and, thus, the prediction error variance is given by

$$\lim_{N\to\infty} \sigma_e^2 = \sigma_n^2 \left\{ \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(1 + \frac{S_h(f)}{\sigma_n^2}\right) df \right) - 1 \right\}. \tag{A.68}$$

A.7 Proof of Equivalency of (5.1) and (5.5) for CM input symbols

In the present appendix, we show that the optimization problems in (5.1), i.e., ML sequence detection, and (5.5), i.e., joint ML detection with MAP parameter estimation, are equivalent in terms of the detected data sequence if the channel fading process and the additive noise are zero-mean jointly proper Gaussian, and if the input symbols have a constant modulus. Therefore, we use the representation given in (5.6) and (5.7), which is equivalent to (5.5). Starting from (5.7) the following holds:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p\left(\mathbf{y}|\mathbf{x}, \mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})\right) p\left(\mathbf{h} = \hat{\mathbf{h}}(\mathbf{x})\right)$$

$$= \arg\max_{\mathbf{x}} \frac{\exp\left(-(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}}(\mathbf{x}))^H \mathbf{R}_n^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}}(\mathbf{x}))\right) \exp\left(-(\hat{\mathbf{h}}(\mathbf{x}))^H \mathbf{R}_h^{-1}\hat{\mathbf{h}}(\mathbf{x})\right)}{\det(\pi\mathbf{R}_n)} \cdot \frac{}{\det(\pi\mathbf{R}_h)}$$

$$= \arg\max_{\mathbf{x}} \left\{ -(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}(\mathbf{x}))^H \mathbf{R}_n^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{h}}(\mathbf{x})) - (\hat{\mathbf{h}}(\mathbf{x}))^H \mathbf{R}_h^{-1}\hat{\mathbf{h}}(\mathbf{x}) \right\}$$

$$\stackrel{(a)}{=} \arg\max_{\mathbf{x}} \left\{ -\left(\mathbf{y} - \mathbf{X}\left(\mathbf{R}_h\mathbf{X}^H(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n)^{-1}\mathbf{y}\right)\right)^H \mathbf{R}_n^{-1} \right.$$

$$\times \left(\mathbf{y} - \mathbf{X}\left(\mathbf{R}_h\mathbf{X}^H(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n)^{-1}\mathbf{y}\right)\right)$$

$$- \left(\mathbf{R}_h\mathbf{X}^H\left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)^{-1}\mathbf{y}\right)^H \mathbf{R}_h^{-1}$$

$$\left. \times \mathbf{R}_h\mathbf{X}^H\left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)^{-1}\mathbf{y}\right\}$$

$$= \arg\max_{\mathbf{x}} \left\{ \mathbf{y}^H \left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)^{-1}\mathbf{y} \right\}$$

$$\stackrel{(b)}{=} \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \tag{A.69}$$

where for (a) we have used that the MMSE channel estimate $\hat{\mathbf{h}}$, which corresponds to the MAP estimate in (5.6) is given by, cf. (A.107)

$$\hat{\mathbf{h}}(\mathbf{x}) = \mathbf{R}_h\mathbf{X}^H \left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)^{-1}\mathbf{y}. \tag{A.70}$$

The matrix $\mathbf{R}_n$ is the autocorrelation matrix of the additive Gaussian noise, which for the special case of white noise is given by $\mathbf{R}_n = \sigma_n^2\mathbf{I}_N$. Furthermore, (b) holds only in case of constant modulus input symbols as $p(\mathbf{y}|\mathbf{x})$ is given by

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\det\left(\pi\left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)\right)} \exp\left(\mathbf{y}^H \left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \mathbf{R}_n\right)^{-1}\mathbf{y}\right) \quad (A.71)$$

where the determinant in the denominator is in general not independent of $\mathbf{x}$. However, for the special case of constant modulus input symbols the determinant is given by $\det(\sigma_x^2\mathbf{R}_h + \mathbf{R}_n)$ and is independent of $\mathbf{x}$. In conclusion, we have shown that the optimization problem in (5.5), i.e., the first line of (A.69), and the optimization problem in (5.1) corresponding to the last line of (A.69) are equal for the special case of constant modulus input symbols.

A.8 Proof of Monotonicity of (6.58)

In this appendix, we prove that the RHS of (6.58), i.e., the argument of the expectation operation, monotonically decreases with an increasing interpolation error variance $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$ on the interval $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}) \in [0, \sigma_h^2]$.

Proof. The argument of the expectation operation at the RHS of (6.58) can be rewritten as follows

$$\int_{z=0}^{\infty} \log\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2(\sigma_h^2 - \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0}))z\right)e^{-z}dz$$

$$- \int_{z=0}^{\infty} \log\left(\sigma_n^2 + \sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})\sigma_x^2 z\right)e^{-z}dz$$

$$= \int_{z=0}^{\infty} \log\left(1 + \frac{\sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_n^2} + \left(\frac{\sigma_x^2\sigma_h^2}{\sigma_n^2} - \frac{\sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_n^2}\right)z\right)e^{-z}dz$$

$$- \int_{z=0}^{\infty} \log\left(1 + \frac{\sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_n^2}z\right)e^{-z}dz$$

$$= \int_{z=0}^{\infty} \log\left(1 + c + (\rho - c)z\right)e^{-z}dz - \int_{z=0}^{\infty} \log\left(1 + cz\right)e^{-z}dz \quad (A.72)$$

where we use (2.18) and define

$$c = \frac{\sigma_x^2\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_n^2} \quad (A.73)$$

with $c \in [0, \rho]$. We define the first and the second integral in (A.72) as $g_1(c)$ and $g_2(c)$

$$g_1(c) = \int_{z=0}^{\infty} \log\left(1 + c + (\rho - c)z\right)e^{-z}dz$$

$$g_2(c) = \int_{z=0}^{\infty} \log\left(1 + cz\right)e^{-z}dz. \quad (A.74)$$

Based on the monotonicity and curvature of $g_1(c)$ and $g_2(c)$ with respect to c, we prove the monotonicity of (A.72) with respect to c. The first and second derivative of $g_1(c)$ and $g_2(c)$ are given by

$$\frac{\partial g_1(c)}{\partial c} = \int_{z=0}^{\infty} \frac{1-z}{1+c+(\rho-c)z} e^{-z} dz \geq 0 \tag{A.75}$$

$$\frac{\partial^2 g_1(c)}{\partial c^2} = \int_{z=0}^{\infty} \frac{-(1-z)^2}{(1+c+(\rho-c)z)^2} e^{-z} dz \leq 0 \tag{A.76}$$

$$\frac{\partial g_2(c)}{\partial c} = \int_{z=0}^{\infty} \frac{z}{1+cz} e^{-z} dz \geq 0 \tag{A.77}$$

$$\frac{\partial^2 g_2(c)}{\partial c^2} = \int_{z=0}^{\infty} \frac{-z^2}{(1+cz)^2} e^{-z} dz \leq 0. \tag{A.78}$$

The inequality in (A.75) is shown in Appendix A.9. Hence, $g_1(c)$ as well as $g_2(c)$ monotonically increase with c and are concave. Therefore, if we can prove the following inequality, we have shown that (A.72) monotonically decreases with $\sigma_{e_{\mathrm{int}}}^2(\mathbf{x}_{\backslash 0})$,

$$g_1(c=\rho) - g_1(c=0) \leq g_2(c=\rho) - g_2(c=0). \tag{A.79}$$

The LHS and the RHS of (A.79) are given by

$$g_1(c=\rho) - g_1(c=0) = \log(1+\rho) - \int_{z=0}^{\infty} \log(1+\rho z)e^{-z}dz \tag{A.80}$$

$$g_2(c=\rho) - g_2(c=0) = \int_{z=0}^{\infty} \log(1+\rho z)e^{-z}dz. \tag{A.81}$$

Introducing (A.80) and (A.81) into (A.79) yields

$$\log(1+\rho) - \int_{z=0}^{\infty} \log(1+\rho z)e^{-z}dz \leq \int_{z=0}^{\infty} \log(1+\rho z)e^{-z}dz$$

$$\Leftrightarrow \quad \log(1+\rho) \leq 2\int_{z=0}^{\infty} \log(1+\rho z)e^{-z}dz$$

$$\Leftrightarrow \quad 0 \leq \int_{z=0}^{\infty} \left\{ \log\left((1+\rho z)^2\right) - \log(1+\rho) \right\} e^{-z}dz = s(\rho). \tag{A.82}$$

To prove (A.82) for $\rho \geq 0$, we show that $s(\rho)$, i.e., the RHS of (A.82) monotonically increases with ρ and use that $s(\rho=0) = 0$. The derivative of $s(\rho)$ with respect to ρ is given by

$$\frac{\partial s(\rho)}{\partial \rho} = \int_{z=0}^{\infty} \left\{ \frac{2z}{1+\rho z} - \frac{1}{1+\rho} \right\} e^{-z}dz. \tag{A.83}$$

To prove that $s(\rho)$ monotonically increases with ρ we show that the RHS of (A.83) is nonnegative, i.e., that the following inequality holds:

$$\int_{z=0}^{\infty} \left\{ \frac{2z}{1+\rho z} - \frac{1}{1+\rho} \right\} e^{-z} dz \geq 0$$

$$\Leftrightarrow \int_{z=0}^{\infty} \frac{2z + \rho z - 1}{1+\rho z} e^{-z} dz \geq 0$$

$$\Leftrightarrow 1 - 2 \int_{z=0}^{\infty} \frac{1-z}{1+\rho z} e^{-z} dz \geq 0$$

$$\Leftrightarrow \frac{1}{\rho} \left[\int_{z=0}^{\infty} \frac{1+\rho}{1+\rho z} e^{-z} dz - 1 \right] \leq \frac{1}{2}$$

$$\Leftrightarrow \int_{z=0}^{\infty} \frac{1+\rho}{1+\rho z} e^{-z} dz \leq \frac{\rho}{2} + 1. \qquad (A.84)$$

The integrand of the integral on the LHS of (A.84) is given by $f(z)e^{-z}$ with

$$f(z) = \frac{1+\rho}{1+\rho z}. \qquad (A.85)$$

To show that (A.84) holds we upper-bound $f(z)$. The function $f(z)$ monotonically decreases and is convex in z, as

$$\frac{\partial f(z)}{\partial z} = -\frac{(1+\rho)\rho}{(\rho z + 1)^2} \leq 0 \qquad (A.86)$$

$$\frac{\partial^2 f(z)}{\partial z^2} = \frac{2(1+\rho)\rho^2}{(\rho z + 1)^3} \geq 0 \quad \forall z \in [0, \infty[. \qquad (A.87)$$

As

$$f(0) = 1 + \rho \qquad (A.88)$$
$$f(1) = 1 \qquad (A.89)$$

we can upper-bound $f(z)$ by a secant between 0 and 1 and by $f(1)$ for $z > 1$, i.e., by the function

$$f_1(z) = \begin{cases} -\rho z + 1 + \rho \text{ for } 0 \leq z \leq 1 \\ 1 \qquad\qquad \text{ for } z > 1 \end{cases}. \qquad (A.90)$$

Hence, (A.84) holds if the following inequality holds:

$$\int_{z=0}^{1} (-\rho z + 1 + \rho) e^{-z} dz + \int_{z=1}^{\infty} e^{-z} dz \leq \frac{\rho}{2} + 1$$

$$\Leftrightarrow \frac{\rho}{e} + 1 \leq \frac{\rho}{2} + 1 \qquad (A.91)$$

which obviously holds as $e \approx 2.718$. Thus, $s(\rho)$ monotonically increases with ρ. Hence, we have shown that (A.82) holds and, therefore, (A.79) is shown. This concludes the proof that (A.72) monotonically decreases with c for $c \in [0, \rho]$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

A.9 Proof of Inequality (A.75)

In this appendix, we prove the following inequality for $c \in [0, \rho]$ and $\rho \in [0, \infty)$

$$\int_{z=0}^{\infty} \frac{1-z}{1+c+(\rho-c)z} e^{-z} dz \geq 0. \tag{A.92}$$

Proof. The LHS of (A.92) can be rewritten as

$$\int_{z=0}^{\infty} \frac{1-z}{1+c+(\rho-c)z} e^{-z} dz$$

$$= \frac{1}{(\rho-c)^2} \left((\rho+1) \exp\left(\frac{1+c}{\rho-c}\right) E_1\left(\frac{1+c}{\rho-c}\right) + c - \rho \right) \tag{A.93}$$

where $E_n(x)$ is the *En*-function defined as follows

$$E_n(x) = \int_1^{\infty} \frac{e^{-xt}}{t^n} dt. \tag{A.94}$$

We have to show that

$$\left((\rho+1) \exp\left(\frac{1+c}{\rho-c}\right) E_1\left(\frac{1+c}{\rho-c}\right) + c - \rho \right) \geq 0$$

$$\Leftrightarrow \exp\left(\frac{1+c}{\rho-c}\right) E_1\left(\frac{1+c}{\rho-c}\right) \geq \frac{\rho-c}{\rho+1}. \tag{A.95}$$

The integral expression in (A.94) can be expressed by the following asymptotic expansion [2, 5.1.51]

$$E_n(x) = \frac{e^{-x}}{x} \left[1 - \frac{n}{x} + \frac{n(n+1)}{x^2} - \frac{n(n+1)(n+2)}{x^3} + \ldots \right]. \tag{A.96}$$

Using (A.96) for the case $n = 1$ and introducing this into the LHS of (A.95), the inequality (A.95) is equivalent to

$$(A.95) \Leftrightarrow \frac{\rho-c}{1+c} \sum_{k=0}^{\infty} k!(-1)^k \left(\frac{\rho-c}{1+c}\right)^k \geq \frac{\rho-c}{\rho+1}$$

$$\Leftrightarrow \sum_{k=0}^{\infty} k!(-1)^k \left(\frac{\rho-c}{1+c}\right)^k \geq \frac{1+c}{1+\rho}$$

$$\stackrel{(a)}{\Leftarrow} \sum_{k=0}^{\infty} k!(-1)^k \left(\frac{\rho-c}{1+c}\right)^k \geq 1 \tag{A.97}$$

where for (a) we used that $c \leq \rho$. Furthermore, we now use the substitution

$$a = \frac{\rho - c}{1 + c} \tag{A.98}$$

where due to the constraints on ρ and c it follows that $a \geq 0$. With (A.98) (A.97) is equivalent to

$$(A.97) \Leftrightarrow \sum_{k=0}^{\infty} k!(-1)^k a^k \geq 1$$

$$\Leftrightarrow \sum_{k=1}^{\infty} k!(-1)^k a^k \geq 0$$

$$\Leftrightarrow \sum_{k=1}^{\infty} (2k)! a^{2k} - \sum_{k=1}^{\infty} (2k-1)! a^{2k-1} \geq 0$$

$$\Leftrightarrow \sum_{k=1}^{\infty} (2k-1)! a^{2k-1} \{2ka - 1\} \geq 0 \tag{A.99}$$

which is obviously true, as for all $a > 0$ the factor $(2ka - 1)$ in (A.99) is positive from a specific k on. This specific k depends on the value of a. Furthermore, infinitely many of these positive terms which are weighted by increasing prefactors due to the factorial are summed up. For the case $a = 0$, the LHS of (A.99) is 0 and thus the inequality (A.99) also holds, which concludes the proof of (A.92). $\qquad\square$

A.10 Comparison of Interpolation and Prediction Error Variance

In this section, we compare the interpolation error variance $\sigma^2_{e_{\text{int,CM},\infty}}$, see (6.75), and the prediction error variance $\sigma^2_{e_{\text{pred,CM},\infty}}|_{\alpha=1}$, see (4.53). Both of them are calculated under the assumption of the transmit symbols being constant modulus with $|x_k|^2 = \sigma_x^2$.

From intuition, it is clear that the interpolation error variance $\sigma^2_{e_{\text{int,CM},\infty}}$ is smaller than the prediction error variance $\sigma^2_{e_{\text{pred,CM},\infty}}|_{\alpha=1}$, as also the knowledge of future observations is used for the estimation. This is a well known result that, e.g., has been shown in [90]. In the following, we also show this relation based on a different proof to the one in [90].

The prediction error variance upper-bounds the interpolation error variance as the following derivation shows:

$$
\begin{aligned}
\sigma^2_{e_{\text{int}},\text{CM},\infty} &= \frac{\sigma^2_n}{\sigma^2_x} \left[\left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\sigma^2_x}{\sigma^2_n} S_h(f) + 1 \right]^{-1} df \right\}^{-1} - 1 \right] \\
&= \frac{\sigma^2_n}{\sigma^2_x} \left[\exp\left(\log\left(\left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\sigma^2_x}{\sigma^2_n} S_h(f) + 1 \right]^{-1} df \right\}^{-1} \right) \right) - 1 \right] \\
&= \frac{\sigma^2_n}{\sigma^2_x} \left[\exp\left(-\log\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\sigma^2_x}{\sigma^2_n} S_h(f) + 1 \right]^{-1} df \right) \right) - 1 \right] \\
&\overset{(a)}{\leq} \frac{\sigma^2_n}{\sigma^2_x} \left[\exp\left(-\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left\{ \left[\frac{\sigma^2_x}{\sigma^2_n} S_h(f) + 1 \right]^{-1} \right\} df \right) - 1 \right] \\
&= \frac{\sigma^2_n}{\sigma^2_x} \left[\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left[\frac{\sigma^2_x}{\sigma^2_n} S_h(f) + 1 \right] df \right) - 1 \right] = \sigma^2_{e_{\text{pred}},\text{CM},\infty} \Big|_{\alpha=1}
\end{aligned}
$$

$$(A.100)$$

where (a) follows from Jensen's inequality.

A.11 Proof for Minimization of $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ in (7.23) by CM Modulation

In this appendix, we will show that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ in (7.23), which depends on the distribution of the data symbols contained in $\mathbf{x}_D$, is minimized for constant modulus input symbols among all distributions of the data symbols with an maximum average power of σ^2_x.

The MAP channel estimate based on pilot and perfectly known data symbols is given by

$$
\begin{aligned}
\hat{\mathbf{h}}_{\text{joint}} &= \arg\max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}, \mathbf{x}) \\
&= \arg\max_{\mathbf{h}} p(\mathbf{y}|\mathbf{h}, \mathbf{x}) p(\mathbf{h}) \\
&= \arg\max_{\mathbf{h}} \{\log(p(\mathbf{y}|\mathbf{h}, \mathbf{x})) + \log(p(\mathbf{h}))\}
\end{aligned}
$$

$$(A.101)$$

with

$$
p(\mathbf{y}|\mathbf{h}, \mathbf{x}) = \frac{1}{\pi^N \sigma^{2N}_n} \exp\left(-\frac{|\mathbf{y} - \mathbf{X}\mathbf{h}|^2}{\sigma^2_n} \right)
$$

$$(A.102)$$

$$
p(\mathbf{h}) = \frac{1}{\pi^N \det(\mathbf{R}_h)} \exp\left(-\mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h} \right).
$$

$$(A.103)$$

Thus, (A.101) becomes

$$\hat{\mathbf{h}}_{\text{joint}} = \arg \max_{\mathbf{h}} \left\{ -\frac{1}{\sigma_n^2} |\mathbf{y} - \mathbf{X}\mathbf{h}|^2 - \mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h} \right\}. \tag{A.104}$$

Differentiating the argument of the maximum operation at the RHS of (A.104) with respect to $\mathbf{h}$ and setting the result equal to zero yields

$$-\frac{1}{\sigma_n^2} \left\{ -\mathbf{X}^H \mathbf{y} + \mathbf{X}^H \mathbf{X}\mathbf{h} \right\} - \mathbf{R}_h^{-1} \mathbf{h} = \mathbf{0} \tag{A.105}$$

and, thus,[1]

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 \mathbf{X}^{-1} \mathbf{X}^{H^{-1}} \right)^{-1} \mathbf{X}^{-1} \mathbf{y}. \tag{A.107}$$

With (A.107) the channel estimation error correlation matrix $\mathbf{R}_{e_{\text{joint}}}$ is given by

$$\mathbf{R}_{e_{\text{joint}}} = \mathrm{E} \left[\left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}} \right) \left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}} \right)^H \middle| \mathbf{x} \right]$$
$$= \mathbf{R}_h - \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h. \tag{A.108}$$

Thus, the differential entropy $h(\mathbf{e}_{\text{joint}} | \mathbf{x}_D, \mathbf{x}_P)$ becomes

$$h(\mathbf{e}_{\text{joint}} | \mathbf{x}_D, \mathbf{x}_P) = \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\pi e \mathbf{R}_{e_{\text{joint}}} \right) \right]$$
$$= \log \left((\pi e)^N \det(\mathbf{R}_h) \right) + \mathrm{E}_{\mathbf{x}} \left[\log \det \left(\mathbf{I}_N - \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h \right) \right]. \tag{A.109}$$

The argument of the expectation operation in the last summand on the RHS of (A.109) can be rewritten as

$$\log \det \left(\mathbf{I}_N - \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h \right)$$
$$= \log \det \left(\mathbf{I}_N - \left(\mathbf{I}_N + \mathbf{R}_h^{-1} \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \right)$$
$$\overset{(a)}{=} \log \det \left(\mathbf{I}_N - \left[\mathbf{I}_N - \left(\frac{1}{\sigma_n^2} \mathbf{R}_h + (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} (\mathbf{X}^H \mathbf{X})^{-1} \right] \right)$$
$$= -\log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathbf{X}^H \mathbf{X} + \mathbf{I}_N \right) \tag{A.110}$$

---

[1] Note that the inverse of $\mathbf{X}$ in (A.107) does not exist, if a diagonal element of the diagonal matrix $\mathbf{X}$ is zero, i.e., one transmit symbol has zero power. However, as the channel estimates can be rewritten as

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \mathbf{X}^H \left(\mathbf{X} \mathbf{R}_h \mathbf{X}^H + \sigma_n^2 \mathbf{I}_N \right)^{-1} \mathbf{y} \tag{A.106}$$

it is obvious that the elements of $\hat{\mathbf{h}}_{\text{joint}}$ are continuous in x_k for all k, and, thus, this does not lead to problems in the following derivation.

where (a) follows from the matrix inversion lemma. Inserting (A.110) into (A.109) yields

$$h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \log\left((\pi e)^N \det(\mathbf{R}_h)\right) - \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathbf{X}^H\mathbf{X} + \mathbf{I}_N\right)\right].$$
(A.111)

As the matrix $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ is diagonal, the product $\mathbf{X}\mathbf{X}^H$ is also diagonal and its diagonal elements are the powers of the individual transmit symbols. In the following we substitute this product by

$$\mathbf{Z} = \mathbf{X}\mathbf{X}^H$$
(A.112)

and $\mathbf{z} = \mathrm{diag}(\mathbf{Z})$ contains the diagonal elements of $\mathbf{Z}$.

The aim of this appendix is to show that the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ corresponding to the entropy in (A.111) is minimized by constant modulus data symbols with the power σ_x^2 among all input distributions fulfilling the maximum average power constraint in (7.1), which is here relaxed to

$$\mathrm{E}\left[\mathbf{x}^H\mathbf{x}\right] = \mathrm{E}\left[\sum_{k=1}^N z_k\right] \leq N\sigma_x^2$$
(A.113)

where the z_k with $k = 1 \ldots N$ are the elements of $\mathbf{z}$. Therefor, in a first step, we study the entropy in (A.111), i.e., a finite transmission length N. Let the set $\mathcal{P}$ be the set containing all input distributions fulfilling the maximum average power constraint in (A.113). Note that this set $\mathcal{P}$ includes the case of having pilot symbols. However, when using pilot symbols, the transmit power of each L-th symbol is fixed to σ_x^2. For the moment, we allow all input distributions contained in $\mathcal{P}$. Later on, we will come back to the special case of using pilot symbols.

We want to find the input vector $\mathbf{z}$ that minimizes (A.111) provided that the average power constraint is fulfilled. Therefor, we first show that the argument of the expectation operation on the RHS of (A.111), i.e.,

$$g(\mathbf{Z}) = \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathbf{Z} + \mathbf{I}_N\right)$$
(A.114)

is concave in $\mathbf{Z}$. To verify the concavity of $g(\mathbf{Z})$, we follow along the lines of [9, Chapter 3.1.5] and consider an arbitrary line $\mathbf{Z} = \bar{\mathbf{Z}} + t\boldsymbol{\Delta}$. Based on this, we define $g(t)$ as

$$
g(t) = \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \left(\bar{\mathbf{Z}} + t\boldsymbol{\Delta} \right) + \mathbf{I}_N \right)
$$

$$
= \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \right) + \log \det \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} + t\boldsymbol{\Delta} \right)
$$

$$
\overset{(a)}{=} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \right) + \log \det \left(\mathbf{Q} + t\boldsymbol{\Delta} \right)
$$

$$
= \log \det \left(\frac{\mathbf{R}_h}{\sigma_n^2} \right) + \log \det \left(\mathbf{Q}^{\frac{H}{2}} \left(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}} \boldsymbol{\Delta} \mathbf{Q}^{-\frac{1}{2}} \right) \mathbf{Q}^{\frac{1}{2}} \right)
$$

$$
= \log \det \left(\frac{\mathbf{R}_h}{\sigma_n^2} \right) + \log \det \left(\mathbf{Q} \right) + \log \det \left(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}} \boldsymbol{\Delta} \mathbf{Q}^{-\frac{1}{2}} \right)
$$

$$
= \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \bar{\mathbf{Z}} + \mathbf{I}_N \right)
$$

$$
+ \log \det \left(\mathbf{I}_N + t \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{H}{2}} \boldsymbol{\Delta} \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{1}{2}} \right)
$$

$$
\overset{(b)}{=} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \bar{\mathbf{Z}} + \mathbf{I}_N \right) + \sum_{k=1}^{N} \log \left(1 + t\lambda_k \right) \tag{A.115}
$$

where for (a) we have used the substitution $\mathbf{Q} = \bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1}$ to simplify notation. Furthermore, the λ_k in (b) are the eigenvalues of $\left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{H}{2}} \boldsymbol{\Delta} \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{1}{2}}$.

Based on (A.115) the derivatives of $g(t)$ with respect to t are given by

$$
\frac{dg(t)}{dt} = \sum_{k=1}^{N} \frac{\lambda_k}{1 + t\lambda_k} \tag{A.116}
$$

$$
\frac{d^2 g(t)}{dt^2} = -\sum_{k=1}^{N} \frac{\lambda_k^2}{\left(1 + t\lambda_k \right)^2}. \tag{A.117}
$$

As the second derivative $\frac{d^2 g(t)}{dt^2}$ is always negative, $g(\mathbf{Z})$ is concave on the set of diagonal matrices $\mathbf{Z}$ with nonnegative diagonal entries.

Based on the concavity of $g(\mathbf{Z})$ with respect to $\mathbf{Z}$ we can lower-bound $h(\mathbf{e}_{\mathrm{joint}} | \mathbf{x}_D, \mathbf{x}_P)$ in (A.111) by using Jensen's inequality as follows, cf. (A.114):

$$
h(\mathbf{e}_{\mathrm{joint}} | \mathbf{x}_D, \mathbf{x}_P) = \log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \mathrm{E}_{\mathbf{z}} \left[g(\mathbf{Z}) \right]
$$

$$
\geq \log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E} \left[\mathbf{Z} \right] + \mathbf{I}_N \right). \tag{A.118}
$$

Recall, that we want to show that constant modulus data symbols with the power σ_x^2 minimize the entropy rate $h'(\mathbf{e}_{\mathrm{joint}} | \mathbf{x}_D, \mathbf{x}_P)$. Therefore, from here on we consider the entropy rate which is given by

$$h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \lim_{N \to \infty} \frac{1}{N} h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$$

$$= \lim_{N \to \infty} \frac{1}{N} \left[\log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \right].$$

$$(A.119)$$

In the next step, we show for which kind of distribution of $\mathbf{z}$ fulfilling the maximum average power constraint in (A.113) the RHS of (A.119) is minimized. I.e., we have to find

$$\lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \qquad (A.120)$$

where the set $\mathcal{P}$ contains all input distributions fulfilling the maximum average power constraint in (A.113).

For the evaluation of (A.120) we substitute the Toeplitz matrix $\mathbf{R}_h$ by an asymptotic equivalent circulant matrix $\mathbf{C}_h$, which is possible, as we are finally interested in the supremum in (A.120) for the case of an infinite transmission length, i.e., $N \to \infty$. In the following, we will formalize the construction of $\mathbf{C}_h$ and show that the following holds

$$\lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right)$$

$$= \lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{C}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \qquad (A.121)$$

Therefore, we express the channel correlation matrix $\mathbf{R}_h$ by its spectral decomposition

$$\mathbf{R}_h = \mathbf{R}_h^{(N)} = \mathbf{U}^{(N)} \mathbf{\Lambda}_h^{(N)} \left(\mathbf{U}^{(N)} \right)^H \qquad (A.122)$$

where we introduced the superscript (N) to indicate the size of the matrices. Furthermore, the matrix $\mathbf{U}^{(N)}$ is unitary and $\mathbf{\Lambda}_h^{(N)} = \mathrm{diag}(\lambda_1^{(N)}, \ldots, \lambda_N^{(N)})$ is diagonal and contains the eigenvalues $\lambda_k^{(N)}$ of $\mathbf{R}_h^{(N)}$.

Analogously to (3.33) in Section 3.3.2, see also [43, Section 4.4, Eq. (4.32)], we construct a circulant matrix $\mathbf{C}_h^{(N)}$ which is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h^{(N)}$. In terms of its spectral decomposition it is given by, cf. (3.39)

$$\mathbf{C}_h^{(N)} = \mathbf{F}^{(N)} \breve{\mathbf{\Lambda}}_h^{(N)} \left(\mathbf{F}^{(N)} \right)^H \qquad (A.123)$$

where the matrix $\mathbf{F}^{(N)}$ is a unitary DFT matrix, i.e., its elements are given by

$$\left[\mathbf{F}^{(N)}\right]_{k,l} = \frac{1}{\sqrt{N}}e^{j2\pi\frac{(k-1)(l-1)}{N}}. \tag{A.124}$$

Furthermore, the matrix $\breve{\boldsymbol{\Lambda}}_h^{(N)}$ is diagonal containing the eigenvalues of $\mathbf{C}_h^{(N)}$. Its elements $\breve{\lambda}_k^{(N)}$ with $k = 1, \dots, N$ are given by, cf. (3.38)

$$\breve{\lambda}_k^{(N)} = \tilde{S}_h\left(\frac{k-1}{N}\right) \tag{A.125}$$

where $\tilde{S}_h(f)$ is the periodic continuation of $S_h(f)$, see (3.34).

By this construction the circulant matrix $\mathbf{C}_h^{(N)}$ is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h^{(N)}$, see [43, Lemma 4.6] and also Section 3.3.2, if

$$\sum_{k=-\infty}^{\infty} |r_h(k)| < \infty \tag{A.126}$$

which is assumed to be fulfilled.

To exploit the asymptotic equivalence of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ for the current problem, we have to show that the matrices in the argument of the log det operation on the LHS and the RHS of (A.121), i.e.,

$$\mathbf{K}_1^{(N)} = \frac{1}{\sigma_n^2}\mathbf{R}_h^{(N)}\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \tag{A.127}$$

$$\mathbf{K}_2^{(N)} = \frac{1}{\sigma_n^2}\mathbf{C}_h^{(N)}\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \tag{A.128}$$

are asymptotically equivalent. However, this has already been shown in Section 3.3.2, cf. (3.48) and (3.49) and substitute $\mathrm{E}\left[\mathbf{Z}\right]$ by $\mathbf{X}^H\mathbf{X}$. As $\mathbf{K}_1^{(N)}$ and $\mathbf{K}_2^{(N)}$ are asymptotically equivalent, with [43, Theorem 2.4] the equality in (A.121) holds. For ease of notation, in the following we omit the use of the superscript (N) for all matrices and eigenvalues.

Based on (A.121) the evaluation of the supremum in (A.120) can be substituted by

$$\lim_{N\to\infty} \frac{1}{N}\sup_{\mathcal{P}}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{C}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right)$$

$$\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N}\sup_{\mathcal{P}}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{F}\breve{\boldsymbol{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right)$$

$$\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N}\sup_{\mathcal{P}}\log\det\left(\frac{1}{\sigma_n^2}\breve{\boldsymbol{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F} + \mathbf{I}_N\right) \tag{A.129}$$

where for (a) we have used (A.123) and (b) is based on (3.25).

As the matrix $\frac{1}{\sigma_n^2}\breve{\boldsymbol{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F} + \mathbf{I}_N$ in the argument of the logarithm on the RHS of (A.129) is positive definite, using Hadamard's inequality we can

upper-bound the argument of the supremum on the RHS of (A.129) as follows

$$\log\det\left(\frac{1}{\sigma_n^2}\check{\boldsymbol{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}+\mathbf{I}_N\right)\leq\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\check{\lambda}_k\left[\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}\right]_{k,k}+1\right)$$

(A.130)

where $\left[\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}\right]_{k,k}$ are the diagonal entries of the matrix $\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}$. Note, this means that distributions of the input sequences $\mathbf{z}$ which lead to the case that the matrix $\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}$ is diagonal maximize the RHS of (A.129). Using (A.130), the RHS of (A.129) is given by

$$\lim_{N\to\infty}\frac{1}{N}\sup_{\mathcal{P}}\log\det\left(\frac{1}{\sigma_n^2}\check{\boldsymbol{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}+\mathbf{I}_N\right)$$

$$=\lim_{N\to\infty}\frac{1}{N}\sup_{\mathcal{P}}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\check{\lambda}_k\left(\frac{1}{N}\sum_{l=1}^{N}\mathrm{E}\left[z_l\right]\right)+1\right)$$

$$=\lim_{N\to\infty}\frac{1}{N}\sup_{\mathcal{P}}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\check{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right)+1\right). \quad\text{(A.131)}$$

It rests to evaluate the supremum on the RHS of (A.131). However, as the logarithm is a monotonically increasing function with the maximum average power constraint in (A.113) the supremum in (A.131) is given by

$$\lim_{N\to\infty}\frac{1}{N}\sup_{\mathcal{P}}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\check{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right)+1\right)$$

$$=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{\sigma_x^2}{\sigma_n^2}\check{\lambda}_k+1\right)$$

$$\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{C}_h+1\right)$$

$$\overset{(b)}{=}\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+1\right) \quad\text{(A.132)}$$

where (a) is based on (A.123) and for (b) we have used the asymptotic equivalence of the circulant matrix $\mathbf{C}_h$ and the Toeplitz matrix $\mathbf{R}_h$.

Now, using (A.121), (A.129), (A.131), and (A.132) the supremum in (A.120) is given by

$$\lim_{N\to\infty}\frac{1}{N}\sup_{\mathcal{P}}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right]+\mathbf{I}_N\right)=\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right).$$

(A.133)

However, this means that the entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ in (A.119) is lower-bounded by

$$
\begin{aligned}
h'&(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P) \\
&= \lim_{N\to\infty} \frac{1}{N} \left[\log\det\left((\pi e)^N \det(\mathbf{R}_h)\right) - \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right) \right] \\
&\geq \lim_{N\to\infty} \frac{1}{N} \left[\log\det\left((\pi e)^N \det(\mathbf{R}_h)\right) - \log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h + \mathbf{I}_N\right) \right] \\
&\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\pi e \mathbf{R}_{e_{\mathrm{joint,CM}}}\right) \tag{A.134}
\end{aligned}
$$

where for (a) we have used (A.109) and (A.110), and where $\mathbf{R}_{e_{\mathrm{joint,CM}}}$ is the estimation error correlation matrix in case all input symbols have a constant modulus with power σ_x^2, i.e., cf. (A.108)

$$
\mathbf{R}_{e_{\mathrm{joint,CM}}} = \mathbf{R}_h - \mathbf{R}_h \left(\mathbf{R}_h + \frac{\sigma_n^2}{\sigma_x^2}\mathbf{I}_N\right)^{-1} \mathbf{R}_h. \tag{A.135}
$$

This mean, that the entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ is minimized for the given maximum average power constraint in (7.1) when all input symbols are constant modulus input symbols with power σ_x^2. Note that this includes the case that each L-th symbol is a pilot symbol with power σ_x^2 and all other symbols are constant modulus data symbols with power σ_x^2.

In conclusion, we have shown that the differential entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ is minimized by constant modulus data symbols with power σ_x^2, i.e.,

$$
\begin{aligned}
h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P) &\geq h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)\big|_{\mathrm{CM}} \\
&= \lim_{N\to\infty} \frac{1}{N} \log\det\left(\pi e \mathbf{R}_{e_{\mathrm{joint,CM}}}\right). \tag{A.136}
\end{aligned}
$$

A.12 Estimation Error Spectra $S_{e_{\mathrm{pil}}}(f)$ and $S_{e_{\mathrm{joint,CM}}}(f)$

In this appendix we will calculate the estimation error spectrum $S_{e_{\mathrm{pil}}}(f)$ of a solely pilot based channel estimation in (7.25), see also (5.15). Afterwards, based on $S_{e_{\mathrm{pil}}}(f)$ we also derive the estimation error spectrum $S_{e_{\mathrm{joint,CM}}}(f)$ in (7.26) for a channel estimation based on all pilot and data symbols for the special case of constant modulus data symbols.

First, we calculate the PSD $S_{e_{\mathrm{pil}}}(f)$ of the channel estimation error in case of a solely pilot based channel estimation. The channel estimation error in the frequency domain is given by

$$E_N(e^{j2\pi f}) = \sum_{k=1}^{N} e_{\text{pil},k} \cdot e^{-j2\pi fk} \tag{A.137}$$

where $e_{\text{pil},k}$ are the elements of the vector $\mathbf{e}_{\text{pil}}$ in (7.13). In the following we are interested in the case $N \to \infty$. As in this case the sum in (A.137) does not exist, in the following we discuss $\lim_{N \to \infty} \frac{1}{N} E_N(e^{j2\pi f})$, which can be expressed as follows

$$\lim_{N \to \infty} \frac{1}{N} E_N(e^{j2\pi f}) \overset{(a)}{=} \lim_{N \to \infty} \frac{1}{N} \sum_{l=1}^{L} E_{N,l}(e^{j2\pi Lf})e^{-j2\pi lf}$$

$$\overset{(b)}{=} \lim_{N \to \infty} \frac{1}{N} \sum_{l=1}^{L} \left[H_{N,l}(e^{j2\pi Lf}) - W_l(e^{j2\pi Lf})\frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] e^{-j2\pi lf}$$

$$\overset{(c)}{=} \lim_{N \to \infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - \sum_{l=1}^{L} W(e^{j2\pi Lf})e^{j2\pi lf}\frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x}e^{-j2\pi lf} \right]$$

$$= \lim_{N \to \infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - L \cdot W(e^{j2\pi Lf})\frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right]$$

$$\overset{(d)}{=} \lim_{N \to \infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - L \cdot W(e^{j2\pi Lf}) \left[H_{N,P}(e^{j2\pi Lf}) + \frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] \right].$$

$$\tag{A.138}$$

For (a) we have used that the estimation error in frequency domain is the sum of the interpolation errors at the individual symbols time instances between the pilot symbols, where the temporal shift yields the phase shift of $2\pi lf$. Here $E_{N,l}(e^{j2\pi Lf})$ is the frequency transform of the estimation error at the symbol position with the distance l to the next pilot symbols, i.e.,

$$E_{N,l}(e^{j2\pi Lf}) = \sum_{k=1}^{\frac{N}{L}} e_{\text{pil},(k-1)L+1+l} \cdot e^{-j2\pi fkL}, \quad \text{for } l = 0\ldots L-1 \tag{A.139}$$

where without loss of generality we assume that N is an integer multiple of L and that the transmit sequence starts with a pilot symbol. Equality (b) results from expressing $E_{N,l}(e^{j2\pi Lf})$ by the difference between the actual channel realization and the estimated channel realization at the different interpolation positions in time domain transferred to frequency domain. Here, without loss of generality, we assume that the pilot symbols are given by σ_x. Furthermore, $W_l(e^{j2\pi Lf})$ is the transfer function of the interpolation filter for the symbols at distance l from the previous pilot symbol. In addition, $Y_{N,P}(e^{j2\pi Lf})$ is the channel output at the pilot symbol time instances transferred to frequency domain. For (c) we have used that the sum of the phase shifted channel realizations in frequency domain at sampling rate $1/L$ corresponds to the frequency domain representation of the fading process at symbol rate. In

addition, we have used that for $N \to \infty$ the interpolation filter transfer function $W_l(e^{j2\pi Lf})$, which is an MMSE interpolation filter, can be expressed as

$$W_l(e^{j2\pi Lf}) = W(e^{j2\pi Lf})e^{j2\pi lf} \tag{A.140}$$

i.e., the interpolation filter transfer functions for the individual time shifts are equal except of a phase shift. Finally, for (d) we have expressed $Y_{N,P}(e^{j2\pi Lf})$ as the sum of the frequency domain representations of the fading process and the additive noise process.

Based on (A.138) the PSD $S_{e_{\text{pil}}}(f)$ is given by

$$
\begin{aligned}
S_{e_{\text{pil}}}(f) &= \lim_{N\to\infty} \frac{1}{N}\mathrm{E}\left[|E_N(e^{j2\pi f})|^2\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\mathrm{E}\Bigg[|H_N(e^{j2\pi f})|^2 - L \cdot H_N(e^{j2\pi f})W^*(e^{j2\pi Lf})H_{N,P}^*(e^{j2\pi Lf}) \\
&\qquad\qquad - L \cdot H_N^*(e^{j2\pi f})W(e^{j2\pi Lf})H_{N,P}(e^{j2\pi Lf}) \\
&\qquad\qquad + L^2|W(e^{j2\pi Lf})|^2\left[|H_{N,P}(e^{j2\pi Lf})|^2 + \left|\frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x}\right|^2\right]\Bigg] \\
&= S_h(e^{j2\pi f}) \\
&\quad - \lim_{N\to\infty} \frac{1}{N}\mathrm{E}\Bigg[L \cdot W^*(e^{j2\pi Lf})\sum_{l=1}^{L} H_{N,l}(e^{j2\pi Lf})e^{-j2\pi lf}H_{N,P}^*(e^{j2\pi Lf}) \\
&\qquad\qquad + L \cdot W(e^{j2\pi Lf})\sum_{l=1}^{L} H_{N,l}^*(e^{j2\pi Lf})e^{j2\pi lf}H_{N,P}(e^{j2\pi Lf})\Bigg] \\
&\quad + L^2|W(e^{j2\pi Lf})|^2\left[\frac{1}{L}S_h(e^{j2\pi Lf}) + \frac{1}{L}\frac{\sigma_n^2}{\sigma_x^2}\right] \\
&= S_h(e^{j2\pi f}) - L \cdot W^*(e^{j2\pi Lf})\sum_{l=1}^{L}\frac{1}{L}S_h(e^{j2\pi Lf}) \\
&\quad - L \cdot W(e^{j2\pi Lf})\sum_{l=1}^{L}\frac{1}{L}S_h^*(e^{j2\pi Lf}) + L|W(e^{j2\pi Lf})|^2\left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}\right] \\
&\overset{(a)}{=} S_h(e^{j2\pi f}) - 2L \cdot W(e^{j2\pi Lf})S_h(e^{j2\pi Lf}) \\
&\quad + L|W(e^{j2\pi Lf})|^2\left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}\right] \tag{A.141}
\end{aligned}
$$

where for (a) we have used that $S_h(f)$ is real and, thus, the MMSE filter $W(e^{j2\pi Lf})$ is also real, see below.

The MMSE filter transfer function $W(e^{j2\pi Lf})$ is given by

$$W(e^{j2\pi Lf}) = \frac{S_h(e^{j2\pi Lf})}{S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}} = \frac{\frac{1}{L}S_h(e^{j2\pi f})}{\frac{1}{L}S_h(e^{j2\pi f}) + \frac{\sigma_n^2}{\sigma_x^2}} \tag{A.142}$$

where we have used that

$$S_h(e^{j2\pi Lf}) = \frac{1}{L}S_h(e^{j2\pi f}). \tag{A.143}$$

Inserting (A.142) into (A.141) yields

$$S_{e_{\mathrm{pil}}}(f) = S_h(e^{j2\pi f}) - 2L\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}S_h(e^{j2\pi Lf})$$

$$+ L\left|\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}\right|^2 \left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}\right]$$

$$\overset{(a)}{=} S_h(e^{j2\pi f}) - 2\cdot\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}S_h(e^{j2\pi f}) + \frac{\left|S_h(e^{j2\pi f})\right|^2}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}$$

$$= \frac{S_h(e^{j2\pi f})L\frac{\sigma_n^2}{\sigma_x^2}}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}$$

$$= \frac{S_h(e^{j2\pi f})}{\frac{\rho}{L}\frac{S_h(e^{j2\pi f})}{\sigma_h^2} + 1}$$

$$\overset{(b)}{=} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} \tag{A.144}$$

where (a) results from (A.143) and for (b) we simplified the notation and substituted $e^{j2\pi f}$ by f to get a consistent notation with (2.3).

The PSD $S_{e_{\mathrm{joint,CM}}}(f)$ is then obviously given by setting $L = 1$ in (A.144), i.e.,

$$S_{e_{\mathrm{joint,CM}}}(f) = \frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2} + 1} \tag{A.145}$$

as all data symbols are assumed to be known and of constant modulus with power σ_x^2, cf. (7.24).

A.13 Proof of Inequality (9.53)

In this appendix, we recall the proof of the inequality (9.53) given in [33] for completeness of presentation.

The goal is to derive an upper bound on the term

$$\frac{1}{T} \int_{\tilde{f}=-\frac{1}{2}}^{\frac{1}{2}} \log \det \left(\frac{PT}{\sigma_n^2 K} \mathbf{S}_h(\tilde{f}) + \mathbf{I}_K \right) d\tilde{f} \tag{A.146}$$

that can be easily evaluated and becomes tight for a large bandwidth, i.e., $K \to \infty$. In order to show tightness in the case of an infinitely large matrix $\mathbf{S}_h(\tilde{f})$ the asymptotic equivalence of Toeplitz and circulant matrices [43] is used. Therefore, (A.146) is upper-bounded by substituting the Toeplitz matrix $\mathbf{S}_h(\tilde{f})$ by its asymptotic equivalent circulant matrix. The advantage in the evaluation of the determinant of a circulant matrix lies in the fact that its eigenvalues can be computed efficiently by using the discrete Fourier transformation.

Let $\mathbf{T}$ be a Hermitian Toeplitz matrix of dimension $K \times K$. In addition, let $\mathbf{F}$ be the $K \times K$ DFT matrix, i.e., a matrix with the elements

$$[\mathbf{F}]_{k,l} = \frac{1}{\sqrt{K}} \exp \left(j2\pi \frac{(k-1)(l-1)}{K} \right). \tag{A.147}$$

By identifying the Hermitian Toeplitz matrix $\mathbf{T}$ with $\frac{PT}{\sigma_n^2} \mathbf{S}_h(\tilde{f})$, an upper bound on the following $\log \det$ expression in (A.146) is required

$$\log \det \left(\mathbf{I}_K + \frac{1}{K} \mathbf{T} \right) \overset{(a)}{=} \log \det \left(\mathbf{I}_K + \frac{1}{K} \mathbf{F}^H \mathbf{T} \mathbf{F} \right)$$

$$\overset{(b)}{\leq} \log \det \left(\mathbf{I}_K + \frac{1}{K} \mathbf{D} \right)$$

$$\overset{(c)}{=} \log \det \left(\mathbf{I}_K + \frac{1}{K} \mathbf{F} \mathbf{D} \mathbf{F}^H \right) \tag{A.148}$$

where $\mathbf{D}$ is a diagonal matrix containing the diagonal entries of $\mathbf{F}^H \mathbf{T} \mathbf{F}$ on its diagonal. For (a) and (c) the fact that the DFT matrix $\mathbf{F}$ is unitary is used, and (b) corresponds to Hadamard's inequality. Obviously, by construction $\mathbf{T}$ and $\mathbf{F} \mathbf{D} \mathbf{F}^H$ are asymptotically equivalent, thus the difference of the LHS and the RHS of the inequality (A.148) disappears for $K \to \infty$.

To apply the inequality (A.148) to the upper-bounding of (A.146), the diagonal entries of $\frac{PT}{\sigma_n^2} \mathbf{F}^H \mathbf{S}_h(\tilde{f}) \mathbf{F}$, which can be identified with $\mathbf{F}^H \mathbf{T} \mathbf{F}$, have to be evaluated. The entries of the Hermitian Toeplitz matrix $\mathbf{S}_h(\tilde{f})$ are given by

$$\mathbf{S}_h(\tilde{f}) = \begin{pmatrix} c_0(\tilde{f}) & c_{-1}(\tilde{f}) & \dots & c_{-(K-1)}(\tilde{f}) \\ c_1(\tilde{f}) & c_0(\tilde{f}) & \dots & c_{-(K-2)}(\tilde{f}) \\ \vdots & \vdots & \ddots & \vdots \\ c_{K-1}(\tilde{f}) & c_{K-2}(\tilde{f}) & \dots & c_0(\tilde{f}) \end{pmatrix} \tag{A.149}$$

with $c_{-k}(\tilde{f}) = c_k^*(\tilde{f})$ and with

$$
\begin{aligned}
c_k(\tilde{f}) &= \sum_{n=-\infty}^{\infty} r_{\mathbb{H}}[n,k] e^{-j2\pi n \tilde{f}} \\
&\overset{(a)}{=} \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_{\tau} S_{\mathbb{H}}\left(\frac{\tilde{f}-n}{T}, \tau\right) e^{-j2\pi kF\tau} d\tau \\
&\overset{(b)}{=} \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_{-\tau_0}^{\tau_0} S_{\mathbb{H}}\left(\frac{\tilde{f}-n}{T}, \tau\right) e^{-j2\pi kF\tau} d\tau
\end{aligned}
\tag{A.150}
$$

where (a) follows from (9.7) and the Poisson summation formula, and (b) uses the fact that the support of the spectrum is limited to $[-\tau_0, \tau_0]$ in the delay domain. Thus, the i-th element on the main diagonal of $\mathbf{F}^H \mathbf{S}_h(\tilde{f})\mathbf{F}$, which are denoted as $d_i(\tilde{f})$ can be expressed by

$$
\begin{aligned}
d_i(\tilde{f}) &= \frac{1}{K} \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} c_{q-p}(\tilde{f}) e^{-j2\pi \frac{i(q-p)}{K}} \\
&= \frac{1}{K} \sum_{k=-(K-1)}^{K-1} (K-|k|) c_k(\tilde{f}) e^{-j2\pi \frac{ik}{K}} \\
&= \mathrm{Re}\left\{ \frac{2}{K} \sum_{k=0}^{K-1} (K-k) c_k(\tilde{f}) e^{-j2\pi \frac{ik}{K}} \right\} - c_0(\tilde{f})
\end{aligned}
\tag{A.151}
$$

where the substitution $k = q - p$ and the relation $c_{-k}(\tilde{f}) = c_k^*(\tilde{f})$ has been used.

With (A.151) the following upper bound on (A.146) can be given

$$
\begin{aligned}
\frac{1}{T} \int_{\tilde{f}=-\frac{1}{2}}^{\frac{1}{2}} &\log \det\left(\frac{PT}{\sigma_n^2 K} \mathbf{S}_h(\tilde{f}) + \mathbf{I}_K \right) d\tilde{f} \\
&\leq \frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i=0}^{K-1} \log\left(1 + \frac{PT}{\sigma_n^2 K} d_i(\tilde{f})\right) d\tilde{f} \\
&\overset{(a)}{=} \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \sum_{i=0}^{K-1} \log\left(1 + \frac{PT}{\sigma_n^2 K} d_i(fT)\right) df \\
&\overset{(b)}{=} \int_{-f_0}^{f_0} \sum_{i=0}^{K-1} \log\left(1 + \frac{PT}{\sigma_n^2 K} d_i(fT)\right) df
\end{aligned}
\tag{A.152}
$$

where for (a) the normalization $f = \tilde{f}/T$ has been used and (b) follows from the support of the spectrum $S_{\mathbb{H}}(f, \tau)$ which is limited to $[-f_0, f_0]$ in the Doppler domain.

For an easy to evaluate upper bound, the dependency of $d_i(fT)$ on f still is an obstacle. This dependency can be removed as follows by further upper-bounding with the help of Jensen's inequality

$$\int_{-f_0}^{f_0} \sum_{i=0}^{K-1} \log\left(1 + \frac{PT}{\sigma_n^2 K} d_i(fT)\right) df \leq 2f_0 \sum_{i=0}^{K-1} \log\left(1 + \frac{PT}{2f_0\sigma_n^2 K} \int_{-f_0}^{f_0} d_i(fT) df\right)$$

$$= 2f_0 \sum_{i=0}^{K-1} \log\left(1 + \frac{P}{2f_0\sigma_n^2 K} d_i\right)$$

(A.153)

with the definition

$$d_i = T \int_{-f_0}^{f_0} d_i(fT) df.$$

(A.154)

For the evaluation of this integral with (A.151) the evaluation of the following expression is needed

$$T \int_{-f_0}^{f_0} c_k(fT) df = \sum_{n=-\infty}^{\infty} \int_{-f_0}^{f_0} \int_{-\tau_0}^{\tau_0} S_{\mathbb{H}}\left(f - \frac{n}{T}, \tau\right) e^{-j2\pi k F \tau} d\tau df$$

$$= \int_{-f_0}^{f_0} \int_{-\tau_0}^{\tau_0} S_{\mathbb{H}}(f, \tau) e^{-j2\pi k F \tau} d\tau df$$

$$= r_{\mathbb{H}}[0, k]$$

(A.155)

where (A.150) has been used. Thus, it follows that

$$d_i = \text{Re}\left\{\frac{2}{K} \sum_{k=0}^{K-1} (K - k) r_{\mathbb{H}}[0, k] e^{-j2\pi \frac{ik}{K}}\right\} - r_{\mathbb{H}}[0, 0]$$

(A.156)

as stated in (9.54). Thus, with (A.152), (A.153), and (A.156) the proof of (9.53) is complete.

Abbreviations

AWGN	additive white Gaussian noise
BER	bit error rate
BPSK	binary phase-shift keying
CM	constant modulus
CSI	channel state information
cu	channel use
DFT	discrete Fourier transformation
DMC	discrete memoryless channel
EM	expectation maximization
i.d.	identically distributed
i.i.d.	independently identically distributed
ICI	intercarrier interference
IDFT	inverse discrete Fourier transformation
ISI	intersymbol interference
LB	lower bound
LHS	left hand side
LLR	log-likelihood ratio
LMMSE	linear minimum mean-square error
MAP	maximum a posteriori
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum likelihood
MMSE	minimum mean-square error
OFDM	orthogonal frequency division multiplexing
QPSK	quadrature phase-shift keying
PDF	probability density function
PG	proper Gaussian
PSD	power spectral density
PSK	phase-shift keying

RC	raised cosine
RHS	ride hand side
SD	synchronized detection
SNR	signal-to-noise rate
SISO	single-input single-output
UB	upper bound
UWB	ultra-wideband
WSSUS	wide-sense stationary uncorrelated scattering

List of Symbols

General

$a, A, b, B, \ldots$	scalars
$\mathbf{a}, \mathbf{b}, \ldots$	vectors
$\mathbf{A}, \mathbf{B}, \ldots$	matrices
$[\mathbf{A}]_{k,l}$	element on k-th row and l-th column of $\mathbf{A}$
$\mathbf{A}^T$	transpose of $\mathbf{A}$
a^*	complex conjugate of a
$\mathbf{A}^H$	conjugate transpose (Hermitian) of $\mathbf{A}$
$\det(\mathbf{A})$	determinant of $\mathbf{A}$
$\mathrm{trace}(\mathbf{A})$	trace of $\mathbf{A}$
$\mathrm{diag}(\mathbf{a})$	diagonal matrix with $\mathbf{a}$ on its main diagonal
$a \star b$	convolution of a and b
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \odot \mathbf{B}$	Hadamard product of $\mathbf{A}$ and $\mathbf{B}$
$\langle a, b \rangle$	inner product of a and b
$\lfloor a \rfloor$	largest integer equal or smaller than a
$\lceil a \rceil$	smallest integer equal or larger than a
$\log(a)$	logarithm of x to the base e
$\mathrm{sinc}(a)$	$sinc$-function, $\mathrm{sinc}(a) = \frac{\sin(\pi a)}{\pi a}$
$\mathrm{tri}(a)$	triangular function, see (A.40)
$E_n(a)$	generalized exponential integral, see (A.94)
$J_0(\cdot)$	zeroth order Bessel function
$\delta(k), \delta_k$	Dirac delta function, Kronecker delta at k
$|\mathbf{A}|$	weak norm (Hilbert-Schmidt norm) of $\mathbf{A}$, see (3.42)
$\|\mathbf{A}\|$	strong norm (operator norm) of $\mathbf{A}$, see (3.52)
$\mathbb{C}$	complex plane
$\gamma \approx 0.577$	Euler constant

Probability and Statistics

$p(a)$	probability density function
$p(a\|b)$	conditional probability density of a given b
$\mathrm{E}[a]$	expectation of the random variable a
$\mathrm{var}[a]$	variance of the random variable a
$\mathcal{CN}(\mu, \mathbf{C})$	jointly proper Gaussian random distribution with mean μ and covariance $\mathbf{C}$

Communication and Information Theory

$\mathcal{I}(\mathbf{y}; \mathbf{x})$	mutual information of $\mathbf{x}$ and $\mathbf{y}$
$\mathcal{I}'(\mathbf{y}; \mathbf{x})$	mutual information rate of $\mathbf{x}$ and $\mathbf{y}$
$h(\mathbf{y})$	differential entropy of $\mathbf{y}$
$h'(\mathbf{y})$	differential entropy rate of $\mathbf{y}$
$D(g\|h)$	Kullback-Leibler distance between the probability distribution functions g and h

$x_k, \mathbf{x}$	channel input, vector: temporally ordered
$h_k, \mathbf{h}$	fading channel, vector: temporally ordered
$n_k, \mathbf{n}$	additive noise, vector: temporally ordered
$y_k, \mathbf{y}$	channel output, vector: temporally ordered
$\mathbf{X}, \mathbf{H}$	diagonal matrices with $\mathbf{x}$ and $\mathbf{h}$ on the main diagonal
N	transmit sequence length

$\hat{h}_k, \hat{\mathbf{h}}, \hat{h}_{\mathrm{pil},k}, \hat{\mathbf{h}}_{\mathrm{pil}}, \hat{\mathbf{h}}_{\mathrm{joint}}$	channel estimates: general, solely pilot based, joint based on pilot and data symbols
$e_k, e_{\mathrm{pil},k}, \mathbf{e}_{\mathrm{pil}}, \mathbf{e}_{\mathrm{joint}}$	channel estimation error: general, solely pilot based channel estimation, joint channel estimation based on pilot and data symbols
$\check{\hat{\mathbf{h}}}_{\mathrm{pil}}$	artificial channel estimation vector with an uncorrelated estimation error

$\mathbf{x}_D, \mathbf{y}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}$	channel input, channel output, channel estimate: all at data symbol time instances only
$\mathbf{x}_P, \mathbf{y}_P$	channel input and output at pilot symbol time instances

$\mathbf{a}_k^l$	vector containing the time instances from k to l of the process $\{a_k\}$
$\mathbf{a}_{\backslash k}$	$\mathbf{a}$ without the time instant k
P_{peak}	peak power
σ_x^2	maximum average power (SISO flat-fading)
α	scaling factor for average power, $\alpha \in [0,1]$
ρ	(nominal) mean SNR
β	nominal peak-to-average power ratio
ν	pilot-to-average data power ratio
σ_p^2	pilot symbol power (if not equal to average data symbol power)
σ_d^2	average data symbol power (if not equal to pilot symbol power)
$\sigma_h^2, \sigma_{\hat{h}}^2, \sigma_n^2$	variance of the channel fading weight h_k, the channel estimate $\hat{h}_k$, and the additive Gaussian noise
$\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$, $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{k-1}^{-\infty})$, $\sigma_{e_{\text{pred,CM},\infty}}^2$	one-step channel prediction error variance, finite past, infinite past, infinite past and CM transmit symbols
$\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}), \sigma_{e_{\text{int,CM},\infty}}^2$	channel interpolation error variance: general case, for CM data symbols and an infinite observation horizon
$\sigma_{e_{\text{pil}}}^2$	estimation error variance of solely pilot based channel estimates
$\mathbf{R}_h$	channel autocorrelation matrix
$\mathbf{\Lambda}_h$	diagonal matrix with eigenvalues of $\mathbf{R}_h$ on its diagonal
λ_i	eigenvalues of $\mathbf{R}_h$
$\mathbf{C}_h$	circulant matrix being asymptotic equivalent to $\mathbf{R}_h$
$\mathbf{F}^{(N)}, \mathbf{F}$	DFT matrix of size $N \times N$, see (3.40)
$\mathbf{R}_{e_{\text{pil}}}, \mathbf{R}_{e_{\text{joint}}}$, $\mathbf{R}_{e_{\text{joint,CM}}}$	correlation matrices of channel estimation error: solely pilot based channel estimation, joint estimation based on pilot and data symbols, joint estimation in case of CM data symbols
$\mathbf{C}_{e_{\text{pil}}}, \mathbf{C}_{e_{\text{joint,CM}}}$	circulant matrices being asymptotic equivalent to $\mathbf{R}_{e_{\text{pil}}}$ and $\mathbf{R}_{e_{\text{joint,CM}}}$
$\mathbf{\Lambda}_{e_{\text{pil}}}, \mathbf{\Lambda}_{e_{\text{joint,CM}}}$	diagonal matrices containing the eigenvalues of $\mathbf{C}_{e_{\text{pil}}}$ and $\mathbf{C}_{e_{\text{joint,CM}}}$

T_{Sym}	symbol duration
f	normalized Doppler frequency (flat-fading)
f_d	maximum normalized Doppler frequency
$r_h(l)$	channel correlation function
$S_h(f)$	channel power spectral density
$S_{e_{\text{pil}}}(f),$ $S_{e_{\text{joint,CM}}}(f)$	power spectral density of channel estimation error: solely pilot based channel estimation, joint estimation based on pilot and CM data symbols
L	pilot spacing
η	SNR degradation factor when using solely pilot based synchronized detection
$\boldsymbol{\lambda}_I$	vector containing reliability information on the information bits
$\tilde{x}_k, \tilde{\mathbf{X}}$	soft-symbols
$\mathcal{P}$	set of all input PDFs with a maximum average power σ_x^2, see (3.3)
$\mathcal{P}_{\text{i.d.}}$	set of all input PDFs with a maximum average power σ_x^2 and i.d. input symbols, see (3.5)
$\mathcal{P}_{\text{i.d.}}^{\text{peak}}$	set of all input PDFs with a maximum average power σ_x^2, a peak power of P_{peak} and i.d. input symbols, see (3.6)
$\mathcal{P}_{\text{i.d.}}^{\text{peak}}\|\alpha$	set of all input PDFs with an average power $\alpha\sigma_x^2$, a peak power of P_{peak} and i.d. input symbols, see (3.69)
$\mathcal{P}_{\text{i.i.d.}}$	set of all input PDFs with a maximum average power σ_x^2 and i.i.d. input symbols, see (4.1)
$\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}$	set of all input PDFs with a maximum average power σ_x^2, a peak power of P_{peak} and i.i.d. input symbols, see (4.2)
$\mathcal{P}_{\mathbf{x}}, \mathcal{P}_{\mathbf{x}_D}$	set of possible transmit (data) sequences (codebook)

MIMO

n_T, n_R	number of transmit and receive antennas
$\tilde{\mathbf{H}}(k), \tilde{\mathbf{x}}(k), \tilde{\mathbf{y}}(k), \tilde{\mathbf{n}}(k)$	channel fading matrix, channel input, channel output, and additive noise at time instant k
$\bar{\mathbf{h}}_{ij}, \bar{\mathbf{x}}_j, \bar{\mathbf{y}}_i, \bar{\mathbf{n}}_i$	vectors containing temporally ordered channel fading weights for subchannel from transmit antenna j to receive antenna i, channel input of transmit antenna j, channel output at receive antenna i, additive noise at receive antenna i
$\bar{\mathbf{X}}_j$	diagonal matrix with $\bar{\mathbf{x}}_j$ on its diagonal
$\mathbf{h}, \mathbf{y}, \mathbf{n}$	channel vector, output vector, and additive noise vector, resulting from stacking $\bar{\mathbf{h}}_{ij}$ of all subchannels, and $\bar{\mathbf{y}}_j$ and $\bar{\mathbf{n}}_j$ of all receive antennas
$\mathbf{X}$	equals $\mathbf{I}_{n_R} \otimes \left[\bar{\mathbf{X}}_1 \dots \bar{\mathbf{X}}_{n_T} \right]$
$\breve{\mathbf{y}}(k), \breve{\mathbf{n}}(k)$	channel output and additive noise vector of spatially decorrelated channel model
$\breve{\mathbf{x}}(k)$	vector of transmit signals on individual eigenmodes with normalized power
$\breve{\mathbf{H}}_w(k)$	channel fading matrix of spatially decorrelated channel model, elements are i.i.d.
$\hat{\tilde{\mathbf{H}}}_{\mathrm{pil}}(k)$	solely pilot based estimate of channel fading matrix at time instant k
$\check{\hat{\mathbf{h}}}_{\mathrm{pil}}$	artificial channel estimate corresponding to $\hat{\tilde{\mathbf{H}}}_{\mathrm{pil}}(k)$ but with a temporally uncorrelated estimation error
$\mathbf{R}_{\tilde{\mathbf{x}}}$	spatial covariance of transmit symbol vector
$\mathbf{V}$	unitary matrix, columns being the eigenvectors of $\mathbf{R}_{\tilde{\mathbf{x}}}$
$\alpha\mathbf{P}$	diagonal matrix containing the eigenvalues of $\mathbf{R}_{\tilde{\mathbf{x}}}$, i.e., the transmit powers in the transmit eigenmodes
$\tilde{\mathbf{P}}$	normalized power distribution matrix, i.e., $\tilde{\mathbf{P}} = \frac{1}{\sigma_x^2}\mathbf{P}$
p_j	maximum average transmit power on transmit eigenmode j
P	maximum average transmit power over all antennas

$\mathbf{R_{\bar{h}}}$	temporal correlation of channel fading weights on individual subchannel
$\mathbf{R_h}$	channel correlation matrix containing spatial and temporal dimensions
$\tilde{\mathbf{\Theta}}_T, \tilde{\mathbf{\Theta}}_R$	transmit and receive antenna correlation
$\tilde{\mathbf{U}}_T, \tilde{\mathbf{U}}_R$	unitary matrices, columns being eigenvectors of $\tilde{\mathbf{\Theta}}_T$ and $\tilde{\mathbf{\Theta}}_R$
$\tilde{\mathbf{\Phi}}$	diagonal matrix containing the eigenvalues of $\tilde{\mathbf{\Theta}}_T$
ϕ_j	diagonal elements of $\tilde{\mathbf{\Phi}}$
$\tilde{\mathbf{\Psi}}$	diagonal matrix containing the eigenvalues of $\tilde{\mathbf{\Theta}}_R$
ψ_i	diagonal elements of $\tilde{\mathbf{\Psi}}$
$\mathcal{P}_{\mathrm{MIMO}}$	set of all input PDFs with a maximum average power P over all input antennas, see (8.33)
$\mathcal{P}_{\mathrm{MIMO,i.d.}}$	set of all input PDFs with an input covariance matrix being capacity-achieving in the coherent case, a maximum average power P over all input antennas, and temporally i.d. distributed input vectors, see (8.36)
$\mathcal{P}_{\mathrm{MIMO,i.d.}}^{\mathrm{peak}}$	set of all input distributions with an input covariance matrix being capacity-achieving in the coherent case, a maximum average power P and a peak power constrained to P_{peak} over all input antennas, and temporally i.d. distributed input vectors, see (8.37)

Frequency Selective Channel

$\mathbb{H}$	linear channel operator in the Hilbert space of square integrable functions
τ	delay
τ_0	maximum delay
f	Doppler shift
f_0	maximum Doppler shift
$\Delta_{\mathbb{H}}$	channel spread $\Delta_{\mathbb{H}} = 4 f_0 \tau_0$
$r_{\mathbb{H}}(t, \nu)$	time-frequency correlation function
$S_{\mathbb{H}}(f, \tau)$	scattering function
$\mathcal{L}_{\mathbb{H}}(t, \nu)$	time-varying transfer function
$\{g_{n,k}(t)\}$	Weyl-Heisenberg set of eigenfunctions

T	spacing of samples of $\mathcal{L}_{\mathbb{H}}(t,\nu)$ in time domain (OFDM symbol duration)
F	spacing of samples of $\mathcal{L}_{\mathbb{H}}(t,\nu)$ in frequency domain (OFDM subcarrier spacing)
$\rho(\tau)$	power delay profile
ρ_l	sample of the power delay profile (average channel power of the l-th delay path)
K	number of frequency slots (subcarriers per OFDM symbol)
B	bandwidth of transmit signal, i.e., $B = KF$
$h[n,k], x[n,k],$ $y[n,k], n[n,k]$	channel fading weight, channel input, channel output, additive noise at time instant n and frequency slot k
$\tilde{\mathbf{h}}_n, \tilde{\mathbf{x}}_n, \tilde{\mathbf{y}}_n, \tilde{\mathbf{n}}_n$	vectors of channel weights, channel input symbols, channel output observations, additive noise samples at time slot n over all frequency slots
$\tilde{\mathbf{H}}_n$	diagonal matrix with $\tilde{\mathbf{h}}_n$ on its diagonal
$\mathbf{h}, \mathbf{x}, \mathbf{y}, \mathbf{n}$	channel, input, output, additive noise vector containing all time and frequency slots, i.e., stacked vectors $\tilde{\mathbf{h}}_n, \tilde{\mathbf{x}}_n, \tilde{\mathbf{y}}_n, \tilde{\mathbf{n}}_n$
$\mathbf{H}, \mathbf{X}$	diagonal matrices with $\mathbf{h}$ and $\mathbf{x}$ on the main diagonal
$\hat{\tilde{\mathbf{H}}}_{\mathrm{pil},n}$	solely pilot based estimate of $\tilde{\mathbf{H}}_n$
$\check{\tilde{\mathbf{h}}}_{\mathrm{pil}}$	artificial channel estimate corresponding to $\hat{\tilde{\mathbf{H}}}_{\mathrm{pil},n}$ but with an uncorrelated channel estimation error in temporal and frequency domain
$\tilde{\mathbf{i}}_n$	interference vector at time slot n over all frequency slots
σ_i^2	interference power
η_I	SNR degradation factor due to interference
P	maximum average transmit power
P_{peak}/K	peak power constraint per subcarrier

$r_{\mathbb{H}}[n,k]$	channel correlation function over time and frequency slots
$\mathbf{R}_{\tilde{\mathbf{h}}}[n]$	channel correlation matrix over subcarriers, between two time slots with a temporal distance n
$\mathbf{R}_{\bar{\mathbf{h}}}[k]$	channel correlation matrix over time, between two subcarriers with a subcarrier distance k
$\mathbf{R_h}$	channel correlation matrix containing frequency and temporal dimension
$S_h(\tilde{f}, \tilde{\tau})$	two-dimensional PSD of channel fading process
$\tilde{f}$	normalized Doppler shift, i.e., $\tilde{f} = fT$
$\tilde{\tau}$	normalized delay, i.e., $\tilde{\tau} = \tau F$
$\mathbf{C}_{\bar{\mathbf{h}}}[0]$	circulant matrix being asymptotic equivalent to $\mathbf{R}_{\bar{\mathbf{h}}}[0]$
$\lambda_{\bar{\mathbf{h}},n}$	eigenvalues of $\mathbf{C}_{\bar{\mathbf{h}}}[0]$
$\lambda_{\tilde{\mathbf{h}},k}$	eigenvalues of $\mathbf{R}_{\bar{\mathbf{h}}}[0]$
$\mathbf{S}_h(\tilde{f})$	matrix valued power spectral density of the multivariate channel fading process $\{\tilde{\mathbf{h}}_n\}$
L_{Sym}	pilot spacing in time domain
L_{SC}	pilot spacing in subcarrier domain
$\mathcal{P}_{\text{OFDM}}$	set of all input PDFs with a maximum average power of P, see (9.39)
$\mathcal{P}_{\text{OFDM,i.d.}}$	set of all input PDFs with a maximum average power of P and i.d. input symbols on all frequency and time slots, see (9.41)
$\mathcal{P}_{\text{OFDM,i.d.}}^{\text{peak}}$	set of all input PDFs with a maximum average power of P, i.d. input symbols on all frequency and time slots, and a peak power constrained to P_{peak}/K per subcarrier, see (9.42)

Discrete Input Distributions

Q	constellation size, cardinality of modulation scheme
$\mathcal{S}$	set of transmit symbol sequences $\mathbf{x}$ of length N
M	cardinality of input set $\mathcal{S}$
$\mathcal{S}_0$	set of pairwise distinguishable transmit sequences of maximal cardinality
p_k	probability of occurrence of input sequence $\mathbf{x}_k$
$\mathbf{p}$	vector containing all p_k for $k \in \{1, \dots, M\}$
$\mathcal{C}$	set over all input distributions $\mathbf{p}$

References

1. I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz). The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. on Information Theory*, 47(4):1290–1301, May 2001.

2. M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, New York, 9th edition, 1972.

3. J. B. Anderson. *Digital Transmission Engineering, 2nd ed.* Wiley-IEEE Press, New York, NY, U.S.A., 2005.

4. J. Baltersee. *On the capacity of multiple-input multiple-output wireless systems with data-aided channel estimation*. PhD thesis, RWTH Aachen University, Aachen, Germany, Oct. 2002.

5. J. Baltersee, G. Fock, and H. Meyr. Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving. *IEEE Journal on Selected Areas Communications*, 19(12):2358–2368, Dec. 2001.

6. J. Baltersee, G. Fock, and H. Meyr. An information theoretic foundation of synchronized detection. *IEEE Trans. on Communications*, 49(12):2115–2123, Dec. 2001.

7. P. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. on Communications Systems*, 11(4):360–393, Dec. 1963.

8. E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. on Information Theory*, 44(6):2619–2692, Oct. 1998.

9. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, U.S.A., 2004.

10. T. H. Chan, S. Hranilovic, and F. R. Kschischang. Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE Trans. on Information Theory*, 51(6):2073–2088, June 2005.

11. J. Chen and V. V. Veeravalli. Capacity results for block-stationary Gaussian fading channels with a peak power constraint. *IEEE Trans. on Information Theory*, 53(12):4498–4520, Dec. 2007.

12. R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow. On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels. *IEEE Trans. on Information Theory*, 50(12):3390–3396, Dec. 2004.

13. R.-R. Chen, R. Koetter, U. Madhow, and D. Agrawal. Joint noncoherent demodulation and decoding for the block fading channel: a practical framework for approaching Shannon capacity. *IEEE Trans. on Communications*, 51(10):1676–1689, Oct. 2003.

14. F. Cheng and A. Venetsanopoulos. Extension of the Toeplitz theorem to the 2-D case and its application to information theory. In *Proc. IEEE International*

Symposium on Circuits and Systems, pages 2486–2489 vol.3, New Orleans, LA, U.S.A., May 1990.

15. D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber. Multiple-input-multiple-output measurements and modeling in Manhattan. *IEEE Journal on Selected Areas Communications*, 21(3):321–331, Apr. 2003.

16. O. Christensen. *An introduction to Frames and Riesz Bases*. Birkhäuser, Boston, MA, U.S.A., 2003.

17. C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. on Information Theory*, 48(3):637–650, Mar. 2002.

18. G. Colavolpe, G. Ferrari, and R. Raheli. Noncoherent iterative (turbo) decoding. *IEEE Trans. on Communications*, 48(9):1488–1498, Sept. 2000.

19. T. Cover and J. Thomas. *Elements of Information Theory, 2nd edition*. Wiley & Sons, New York, 2006.

20. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

21. X. Deng and A. M. Haimovich. Information rates of time varying Rayleigh fading channels. In *Proc. IEEE International Conference on Communications (ICC)*, volume 1, pages 573–577, Paris, France, June 2004.

22. X. Deng and A. M. Haimovich. Achievable rates over time-varying Rayleigh fading channels. *IEEE Trans. on Communications*, 55(7):1397–1406, July 2007.

23. D. Divsalar and M. K. Simon. Multiple-symbol differential detection of MPSK. *IEEE Trans. on Communications*, 38(3):300–308, Mar. 1990.

24. J. Doob. *Stochastic Processes*. Wiley, New York, 1990.

25. M. Dörpinghaus, G. Ascheid, H. Meyr, and R. Mathar. Optimal PSK signaling over stationary Rayleigh fading channels. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 126–130, Toronto, Canada, July 2008.

26. M. Dörpinghaus, A. Ispas, G. Ascheid, and H. Meyr. On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels. In *Proc. of the 2010 International Zurich Seminar on Communications*, pages 74–77, Zurich, Switzerland, Mar. 2010.

27. M. Dörpinghaus, H. Meyr, and G. Ascheid. The achievable rate of stationary Rayleigh flat-fading channels with IID input symbols. In *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pages 812–817, Taichung, Taiwan, Oct. 2010.

28. M. Dörpinghaus, M. Senst, G. Ascheid, and H. Meyr. On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian input distribution. In *Proc. International Symposium on Information Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.

29. L. Duan and B. Rimoldi. The iterative turbo decoding algorithm has fixed points. *IEEE Trans. on Information Theory*, 47(7):2993–2995, Nov. 2001.

30. G. Durisi, H. Bölcskei, and S. Shamai (Shitz). Capacity of underspread WSSUS fading channels in the wideband regime. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 1500–1504, Seattle, WA, U.S.A., July 2006.

31. G. Durisi, H. Bölcskei, and S. Shamai (Shitz). Capacity of underspread noncoherent WSSUS fading channels under peak signal constraints. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 156–160, Nice, France, June 2007.

32. G. Durisi, V. I. Morgenshtern, and H. Bölcskei. On the sensitivity of noncoherent capacity to the channel model. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

33. G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz). Noncoherent capacity of underspread fading channels. *IEEE Trans. on Information Theory*, 56(1):367–395, Jan. 2010.

34. R. H. Etkin and D. N. C. Tse. Degrees of freedom in some underspread MIMO fading channels. *IEEE Trans. on Information Theory*, 52(4):1576–1608, Apr. 2006.

35. A. Feiten and R. Mathar. Capacity-achieving discrete signaling over additive noise channels. In *Proc. IEEE International Conference on Communications (ICC), Glasgow, U.K.*, pages 5401–5405, June 2007.

36. G. Foschini and M. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications, Springer Netherlands*, 6(3):311–335, Mar. 1998.

37. S. Furrer and D. Dahlhaus. Multiple-antenna signaling over fading channels with estimated channel state information: Performance analysis. *IEEE Trans. on Information Theory*, 53(6):2010–2027, June 2007.

38. R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, NY, U.S.A., 1968.

39. H. Gazzah, P. A. Regalia, and J. P. Delmas. Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification. *IEEE Trans. on Information Theory*, 47(3):1243–1251, Mar. 2001.

40. S. Godtmann. *Iterative Synchronization at Low Signal-to-Noise Ratios for Wireless Communications*. PhD thesis, RWTH Aachen University, Aachen, Germany, Mar. 2009.

41. S. Godtmann, N. Hadaschik, A. Pollok, G. Ascheid, and H. Meyr. Iterative code-aided phase noise synchronization based on the LMMSE criterion. In *Proc. IEEE 8th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Helsinki, Finland*, June 2007.

42. A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas Communications*, 21(5):684–702, June 2003.

43. R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2006.

44. U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications*. Univ. Calif. Press, Berkeley, CA, U.S.A., 1958.

45. D. Guo, S. Shamai (Shitz), and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. on Information Theory*, 51(4):1261–1282, Apr. 2005.

46. D. Guo, S. Shamai (Shitz), and S. Verdú. Proof of entropy power inequalities via MMSE. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 1011–1015, Seattle, WA, U.S.A., July 2006.

47. N. Hadaschik, G. Ascheid, and H. Meyr. Achievable data rate of wideband OFDM with data-aided channel estimation. In *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, Finland, Sept. 2006.

48. B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Trans. on Information Theory*, 49(4):951–963, Apr. 2003.

49. W. He and C. N. Georghiades. Computing the capacity of a MIMO fading channel under PSK signaling. *IEEE Trans. on Information Theory*, 51(5):1794–1803, May 2005.

50. B. M. Hochwald and T. L. Marzetta. Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading. *IEEE Trans. on Information Theory*, 46(2):543–564, Mar. 2000.

51. P. Hoeher and J. Lodge. Turbo DPSK: Iterative differential PSK demodulation and channel decoding. *IEEE Trans. on Communications*, 47(6):837–843, June 1999.

52. R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, Cambridge, U.K., 1985.

53. S. A. Jafar and A. Goldsmith. Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information. *IEEE Trans. Wireless Communications*, 4(3):990–997, May 2005.

54. W. C. Jakes, editor. *Microwave Mobile Communications*. John Wiley & Sons Inc, New York, NY, U.S.A., 1975.

55. N. Jindal and A. Lozano. Optimum pilot overhead in wireless communication: A unified treatment of continuous and block-fading channels. *IEEE Trans. Wireless Communications*, submitted. [Online]. Available: http://arxiv.org/abs/0903.1379.

56. N. Jindal, A. Lozano, and T. Marzetta. What is the value of joint processing of pilots and data in block-fading channels? In *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

57. E. A. Jorswieck and H. Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Trans. Wireless Communications*, 3(5):1543–1553, Sept. 2004.

58. A. Kavcic, J. M. F. Moura, and V. Bhagavatula. Iterative, soft signal processing for digital communications - from the guest editors. *IEEE Signal Processing Magazine*, 21(1):26–108, Jan. 2004.

59. T. Koch and A. Lapidoth. The fading number and degrees of freedom in noncoherent MIMO fading channels: A peace pipe. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 661–665, Adelaide, Australia, Sept. 2005.

60. T. Koch and A. Lapidoth. Multipath channels of bounded capacity. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 6–10, Porto, Portugal, May 2008.

61. T. Koch and A. Lapidoth. Multipath channels of unbounded capacity. In *Proc. IEEE 25th Convention of Electrical and Electronics Engineers in Israel, IEEEI*, pages 640–644, Eilat, Israel, Dec. 2008.

62. T. Koch and A. Lapidoth. On multipath fading channels at high SNR. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 1572–1576, Toronto, ON, Canada, July 2008.

63. T. Koch and A. Lapidoth. On multipath fading channels at high SNR. *IEEE Trans. on Information Theory*, 56(12):5945–5957, Dec. 2010.

64. W. Kozek. *Matched Weyl-Heisenberg expensions of nonstationary environments*. PhD thesis, Vienna University of Technology, Vienna, Austria, Mar. 1997.

65. W. Kozek and A. F. Molisch. Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels. *IEEE Journal on Selected Areas Communications*, 16(8):1579–1589, Oct. 1998.

66. A. Lapidoth. On the high SNR capacity of stationary Gaussian fading channels. In *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, pages 410–419, Monticello, IL, U.S.A., Oct. 2003.

67. A. Lapidoth. On the asymptotic capacity of stationary Gaussian fading channels. *IEEE Trans. on Information Theory*, 51(2):437–446, Feb. 2005.

68. A. Lapidoth and S. M. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. on Information Theory*, 49(10):2426–2467, Oct. 2003.

69. A. Lapidoth and S. M. Moser. The fading number of single-input multiple-output fading channels with memory. *IEEE Trans. on Information Theory*, 52(2):437–453, Feb. 2006.

70. A. Lapidoth and S. Shamai (Shitz). Fading channels: how perfect need "perfect side information" be? *IEEE Trans. on Information Theory*, 48(5):1118–1134, May 2002.

71. Y. Liang and V. V. Veeravalli. Capacity of noncoherent time-selective Rayleigh-fading channels. *IEEE Trans. on Information Theory*, 50(12):3095–3110, Dec. 2004.

72. I. D. Marsland and P. T. Mathiopoulos. Multiple differential detection of parallel concatenated convolutional (turbo) codes in correlated fast Rayleigh fading. *IEEE Journal on Selected Areas Communications*, 16(2):265–275, Feb. 1998.

73. I. D. Marsland and P. T. Mathiopoulos. On the performance of iterative non-coherent detection of coded M-PSK signals. *IEEE Trans. on Communications*, 48(4):588–596, Apr. 2000.

74. G. Maruyama. The harmonic analysis of stationary stochastic processes. *Memoirs of the Faculty of Science, Series A, Mathematics*, 4(1):45–106, 1949.

75. T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. on Information Theory*, 45(1):139–157, Jan. 1999.

76. G. Matz, D. Schafhuber, K. Gröchenig, M. Hartmann, and F. Hlawatsch. Analysis, optimization, and implementation of low-interference wireless multicarrier systems. *IEEE Trans. Wireless Communications*, 6(5):1921–1931, May 2007.

77. M. Médard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Trans. on Information Theory*, 46(3):933–946, May 2000.

78. M. Médard and R. G. Gallager. Bandwidth scaling for fading multipath channels. *IEEE Trans. on Information Theory*, 48(4):840–852, Apr. 2002.

79. H. Meyr, M. Moeneclaey, and S. Fechtel. *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing, 1st ed.* John Wiley & Sons, New York, NY, U.S.A., 1998.

80. M. Miranda and P. Tilli. Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results. *SIAM Journal on Matrix Analysis and Applications*, 21(3):867–881, Feb. 2000.

81. S. M. Moser. The fading number of multiple-input multiple-output fading channels with memory. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 521–525, Nice, France, June 2007.

82. F. D. Neeser and J. L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. on Information Theory*, 39(4):1293–1302, July 1993.

83. N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe. A theoretical framework for soft-information-based synchronization in iterative (Turbo) receivers. *EURASIP Journal on Wireless Communications and Networking*, 2005(2):117–129, 2005.

84. A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge Univ. Press, 2003.

85. M. Peleg and S. Shamai (Shitz). Iterative decoding of coded and interleaved noncoherent multiple symbol detected DPSK. *Electronics Letters*, 33(12):1018–1020, June 1997.

86. M. Peleg, S. Shamai (Shitz), and S. Galan. Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel. *IEE Proc. Communications*, 147(2):87–95, Apr. 2000.

87. R. R. Perera. *On capacity of fading channels with no channel state information*. PhD thesis, The Australian National University, Melbourne, Australia, June 2006.

88. R. R. Perera, T. S. Pollock, and T. D. Abhayapala. Performance of Gaussian distributed input in non-coherent Rayleigh fading MIMO channels. In

Proc. IEEE International Conference on Communications and Signal Processing (ICICS2005), pages 791–795, Bangkok, Thailand, Dec. 2005.

89. K. B. Petersen and M. S. Pedersen. The matrix cookbook, February 2008.

90. B. Picinbono and J. M. Kerilis. Some properties of prediction and interpolation errors. *IEEE Trans. on Acoust., Speech, Signal Process. (1975-1990)*, 36(4):525–531, Apr. 1988.

91. M. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, CA, U.S.A., 1964.

92. D. Porrat, D. N. C. Tse, and S. Nacu. Channel uncertainty in ultra-wideband communication systems. *IEEE Trans. on Information Theory*, 53(1):194–208, Jan. 2007.

93. F. Rusek, A. Lozano, and N. Jindal. Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 300–304, Austin, TX, U.S.A., June 2010.

94. D. Samardzija and N. Mandayam. Pilot-assisted estimation of MIMO fading channel response and achievable data rates. *IEEE Trans. on Signal Processing*, 51(11):2882–2890, Nov. 2003.

95. F. Sanzi, S. Jelting, and J. Speidel. A comparative study of iterative channel estimators for mobile OFDM systems. *IEEE Trans. Wireless Communications*, 2(5):849–859, Sep. 2003.

96. D. Schafhuber. *Wireless OFDM systems: Channel prediction and system capacity*. PhD thesis, Vienna University of Technology, Vienna, Austria, March 2004.

97. D. Schafhuber, H. Bölcskei, and G. Matz. System capacity of wideband OFDM communications over fading channels without channel knowledge. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2004. Correction in Eq. 2 compared to the version published in the proceedings., http://www.nari.ee.ethz.ch/commth/pubs/p/ofdm04.

98. L. Schmitt. *On iterative receiver algorithms for concatenated codes*. PhD thesis, RWTH Aachen University, Aachen, Germany, Jan. 2008.

99. U. G. Schuster. *Wireless Communication over Wideband Channels*. PhD thesis, ETH Zurich, Zurich, Switzerland, 2009.

100. U. G. Schuster and H. Bölcskei. Ultrawideband channel modeling on the basis of information-theoretic criteria. *IEEE Trans. Wireless Communications*, 6(7):2464–2475, July 2007.

101. U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor. Capacity bounds for peak-constrained multiantenna wideband channels. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 1582–1586, Toronto, ON, Canada, July 2008.

102. U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor. Capacity bounds for peak-constrained multiantenna wideband channels. *IEEE Trans. on Communications*, 57(9):2686–2696, Sept. 2009.

103. V. Sethuraman and B. Hajek. Capacity per unit energy of fading channels with a peak constraint. *IEEE Trans. on Information Theory*, 51(9):3102–3120, Sept. 2005.

104. V. Sethuraman and B. Hajek. Low SNR capacity of fading channels with peak and average power constraints. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 689–693, Seattle, WA, U.S.A., July 2006.

105. V. Sethuraman, B. Hajek, and K. Narayanan. Capacity bounds for noncoherent fading channels with a peak constraint. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 515–519, Adelaide, Australia, Sept. 2005.

106. V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth. Low SNR capacity of fading channels – MIMO and delay spread. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 516–520, Nice, France, June 2007.

107. V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth. Low-SNR capacity of noncoherent fading channels. *IEEE Trans. on Information Theory*, 55(4):1555–1574, Apr. 2009.

108. C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and Oct. 1948.

109. D.-S. Shiu, G. Foschini, M. Gans, and J. Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans. on Communications*, 48(3):502–513, Mar. 2000.

110. P. H. Siegel, D. Divsalar, E. Eleftheriou, J. Hagenauer, D. Rowitch, and W. H. Tranter. Guest editorial - The Turbo principle: From theory to practice. *IEEE Journal on Selected Areas Communications*, 19(5):793–799, May 2001.

111. P. H. Siegel, D. Divsalar, E. Elepftheriou, J. Hagenauer, and D. Rowitch. Guest editorial - The Turbo principle: From theory to practice II. *IEEE Journal on Selected Areas Communications*, 19(9):1657–1661, Sept. 2001.

112. B. Sklar. Rayleigh fading channels in mobile digital communication systems part I: Characterization. *IEEE Communications Magazine*, 35(7):90–100, July 1997.

113. D. Slepian. On bandwidth. *Proceedings of the IEEE*, 64(3):292–300, Mar. 1976.

114. S. G. Srinivasan and M. K. Varanasi. Optimal spatial correlations for the noncoherent MIMO Rayleigh fading channel. *IEEE Trans. Wireless Communications*, 6(10):3760–3769, Oct. 2007.

115. S. G. Srinivasan and M. K. Varanasi. Optimal constellations for the low-SNR noncoherent MIMO block Rayleigh-fading channel. *IEEE Trans. on Information Theory*, 55(2):776–796, Feb. 2009.

116. N. M. Steen, G. D. Byrne, and E. M. Gelbard. Gaussian quadrature for the integrals. *Mathematics of Computation*, 23(107):661–671, July 1969.

117. V. G. Subramanian and B. Hajek. Broad-band fading channels: Signal burstiness and capacity. *IEEE Trans. on Information Theory*, 48(4):809–827, Apr. 2002.

118. G. Szegö. Beiträge zur Theorie der Toeplitzschen Formen. *Mathematische Zeitschrift*, 6(3-4):167–202, Sept. 1920.

119. G. Taricco and M. Elia. Capacity of fading channel with no side information. *Electronics Letters*, 33(16):1368–1370, July 1997.

120. I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. on Telecommunications*, 10(6):585–595, Nov./Dec. 1999.

121. I. E. Telatar and D. N. C. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Trans. on Information Theory*, 46(4):1384–1400, July 2000.

122. A. Tulino and S. Verdú. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1):1–182, 2004.

123. A. M. Tulino, A. Lozano, and S. Verdú. Impact of antenna correlation on the capacity of multiantenna channels. *IEEE Trans. on Information Theory*, 51(7):2491–2509, July 2005.

124. A. M. Tulino, A. Lozano, and S. Verdú. Capacity-achieving input covariance for single-user multi-antenna channels. *IEEE Trans. Wireless Communications*, 5(3):662–671, Mar. 2006.

125. M. C. Valenti and B. D. Woerner. Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels. *IEEE Journal on Selected Areas Communications*, 19(9):1697–1705, Sept. 2001.

126. R. Vaughan and J. Bach Andersen. *Channels, Propagation and Antennas for Mobile Communications*. The Institution of Electrical Engineers, London, U.K., 2003.

127. S. Verdú. On channel capacity per unit cost. *IEEE Trans. on Information Theory*, 36(5):1019–1030, Sept. 1990.

128. S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. on Information Theory*, 48(6):1319–1343, June 2002.

129. A. J. Viterbi. On the minimum mean square error resulting from linear filtering of stationary signals in white noise (corresp.). *IEEE Trans. on Information Theory*, 11(4):594–595, Oct. 1965.

130. D. Warrier and U. Madhow. Spectrally efficient noncoherent communication. *IEEE Trans. on Information Theory*, 48(3):651–668, Mar. 2002.

131. H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels. *IEEE Trans. on Information Theory*, 50(8):1665–1686, Aug. 2004.

132. W. Zhang and J. N. Laneman. Benefits of spatial correlation for multi-antenna non-coherent communication over fading channels at low SNR. *IEEE Trans. Wireless Communications*, 6(3):887–896, Mar. 2007.

133. W. Zhang and J. N. Laneman. How good is PSK for peak-limited fading channels in the low-SNR regime? *IEEE Trans. on Information Theory*, 53(1):236–251, Jan. 2007.

134. L. Zheng and D. N. C. Tse. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. on Information Theory*, 48(2):359–383, Feb. 2002.

Index